Improving mathematical abilities by training numerical representations in children: the relation between learning mathematics and numerical cognition

Nuria Ferres Forga

TESI DOCTORAL UPF / 2017

DIRECTOR DE LA TESI

Dr. Luca L. Bonatti

DEPARTAMENT DE TECNOLOGIES DE LA INFORMACIÓ I LES COMUNICAIONS

Universitat Pompeu Fabra Barcelona

Dedico aquest doctorat al meu estimat marit Carles per la confiança que sempre a tingut en mi; i també als meus fills, Ariadna i Abel, que han sigut el motor i la inspiració que m'han calgut per arribar fins aquí. I al meu germà que és un gran professor de matemàtiques. Al meu pare que no ha pogut veure culminar aquest doctorat, però que ell sabia que ho aconseguiria. A la meva mare, per tot. A la meva tieta, també. I als meus nebots perquè són un cel.

Acknowledgements

Thank you very much.

To my family for everything.

To the colleagues and friends from the doctorate for their emotional support.

To my friend Sònia Sas, the general director of Hamelin Laie International School, for providing me participants and school facilities to conduct all the studies of this thesis.

To all the teachers and staff of the Hamelin Laie International School, for their kind collaboration.

To the 710 students aged 7 to 16 who participated in the studies.

To my director of the thesis, Luca Bonatti, for his scientific support.

To Justin Halberda for his scientific support.

To all of them, thanks again.

Abstract

Improving mathematical abilities is important for educational systems and for society overall. We present two training regimes based on numerical representations.

In Study 1, we show that a three-week computer-based quantity discrimination training, focused on enhancing the accuracy of the Approximate Number System (ANS), improved mathematics performance in low-performing 7-to-8-year-old children.

In Study 2, we show that a novel numerical estimation training enhancing mappings between Arabic digits and quantities, improved overall mathematical competence in all children, going beyond the improvements obtained by training ANS.

In Study 3, we show that performance in both trainings correlate, in different extend, with school math marks, but we especially found a consistent and extended relation between the ability of mapping digits to quantities and the school math marks in pupils from 8 to 13 years of age.

Thus, training the precision of the digit-quantity relation may improve mathematical competence, particularly in the first crucial years of exposure to formal mathematics.

Resum

La millora de les habilitats matemàtiques es un objectiu important pels sistemes educatius i per la societat en general. Aquí presentem dos entrenaments basats en representacions numèriques.

A l’estudi 1, mostrem com entrenant la discriminació de quantitats durant tres setmanes amb la intenció de fer més precís el Sistema d’Aproximació Numèric (ANS), millora el rendiment matemàtic del nens de 7-8 anys de baix perfil acadèmic

A l’estudi 2, mostrem com un inèdit entrenament d’estimació numèrica dissenyat per incrementar la precisió en relacionar els dígits Aràbics amb les quantitats que representen, provoca una millora generalitzada de la competència matemàtica en tots els perfils acadèmics en nens de 7 a 8 anys.

A l’estudi 3, mostrem com el rendiment en els dos entrenaments correlaciona amb les notes de matemàtiques a l’escola encara que en diferent mesura, essent l’habilitat de relacionar els dígits amb les quantitats, la que correlaciona amb les notes en més cursos escolars, des dels 8 anys fins als 13.

Així, entrenar la precisió en relacionar els dígits amb quantitats provoca una millora de la competència matemàtica, sobretot en els primers anys crucials d'exposició a les matemàtiques formals.

Preface

Across school years, many children struggle with mathematics affecting their confidence and future career choice, eventually creating a social problem. Society in general, and the educational system in particular, have a responsibility for the generation of this problematic situation. Indeed, several societal factors contribute to it: socioeconomic status, math anxiety, gender discrimination, or teacher biases, among others. Another important factor, we believe, is that the educational systems overestimate 7-8 year olds' comprehension of some basic aspects of mathematical language. An appropriate, evidence-based training of basic number abilities may complement standard school teaching routines, in an effortless and playful way, potentially generating long-lasting benefits in children's mathematical abilities and self-confidence, starting from these crucial founding years of their initiation to mathematical concepts. In the classroom, one typically observes a range of mathematical skill from low-performing to high-performing children. Parents, teachers, and community leaders all share a desire to help children improve in school. Perhaps this desire can be most keenly felt for low-performing children who are struggling alongside their middle- and high-performing peers. The mathematical abilities required in the classroom range from simpler skills, like intuitive estimation and comparison of numbers of items, to more complex abilities such as multiplication. One interest of researchers in cognitive science and education is to understand the range of mathematical content taught in school (e.g., addition, subtraction, multiplication) and the range of basic cognitive abilities that support this content.

I would like to confess that I am very pleased to have contributed to both, the research in numerical cognition and the educational system. The *Digits* computer program that I have developed, for the time being, has met the expectations of helping children improve their mathematical abilities.

I hope you enjoy your reading. Nuria Ferres Forga

Table of contents

1. INTRODUCTION

1.1 Learning Mathematics

Mathematical competence is a fundamental factor in important aspects of life. It correlates with socioeconomic status and success in the professional world (Parsons & Bynner, 2005; Ritchie & Bates, 2013); with health status (Peters, Hart, Tusler, & Fraenkel, 2014); and of course with academic achievements (Geary, Hoard, Nugent, & Bailey, 2013; Jordan, Kaplan, Ramineni, & Locuniak, 2009; Papay, Murnane, & Willett, 2014; Watts, Duncan, Siegler, & Davis-Kean, 2014). It is thus equally fundamental to understand how mathematical skills are acquired and how they can be improved, especially in the critical early years of schooling.

The process of learning mathematics to enhance mathematical competence can be approached from two dimensions. First, social and educational aspects have a particularly important influence. Numerous scientific studies have been dedicated to this dimension. Socioeconomic status plays a role (Ramani & Siegler, 2011; Thien & Ong, 2015; Verdine et al., 2014), but so do gender (J. S. Hyde & Mertz, 2009; Stoet, Bailey, Moore, & Geary, 2016), math anxiety (Jansen et al., 2013; Maloney & Beilock, 2012; M.I. Núñez-Peña, Suárez-Pellicioni, & Bono, 2013; Pletzer, Kronbichler, Nuerk, & Kerschbaum, 2015; Z. Wang et al., 2014), motivation and predisposition to mathematics (Cerda et al., 2015; Simzar, Domina, & Tran, 2016), language development and bilingualism (Moll, Snowling, Göbel, & Hulme, 2015; Spelke & Tsivkin, 2001; Van Rinsveld, Brunner, Landerl, Schiltz, & Ugen, 2015), or teacher effects (Demaray & Elliot, 1998; Desimone & Long, 2010; Tournaki, 2003). More recently, we know that the use of educational software and game-based learning can be added to this already long list (Bugden, DeWind, & Brannon, 2016; Desoete & Praet, 2013; Niederhauser & Stoddart, 2001; Praet & Desoete, 2014).

Second, an extensive scientific literature documents the importance of cognitive processes in the acquisition of mathematical skills, starting from those more directly related to numerical cognition. We will expand on them in the next few sections. To ensure that students can develop their mathematical skills in the most appropriate, enriching and efficient way possible, these two approaches should be more interconnected so that they can benefit each other (Coch & Ansari, 2009; Siegler, 2003),. Furthermore, when we deal with children, this process should also be easy, interesting, and fun as much as possible. But to achieve these goals, a deep understanding of the pillars on which the construction of mathematical competence can be built is necessary.

How can scientific knowledge on numerical cognition help students at school to learn and understand mathematics better? And, at the same time, how can students' numerical performance at school serve to broaden our knowledge of cognition? In this thesis, we have focused our research on this double dilemma: using what we know about numerical cognition to improve mathematical education, and using mathematical performance in educational settings to have a glimpse on some aspects of the underlying cognitive processes. Inspired by the current scientific knowledge on numerical cognition and mathematics education, enriched by my personal experiences and knowledge in the field of education, I have created a numerical computer training regime. I adapted it to children in order to help them learning mathematics at early and crucial years in which this learning takes place. I also explored how it can be also used to test numerical abilities in children and teenagers. I called this game the *Digits* game. Basically, Digits game trains the precision in mapping Arabic digits to the quantities they represent by a numerical estimation task.

To advance our knowledge of the effectiveness of existing methods, while at the same time creating a yardstick for the method I elaborate, I have explored the results of a prolonged training with an existing computer testing program, the *Panamath* game (Halberda, Mazzocco, & Feigenson, 2008), which has been studied over large populations (Halberda, Ly, Wilmer, Naiman, & Germine, 2012), although its effectiveness in prolonged training were unknown. This program measures the approximate number system abilities by means of a quantity discrimination task. We used it here as a test for students, but also as a training method for children. In this way, at the same time, we could test aspects of the program which were unknown and compare our novel method with a more established, scientifically motivated, computer-based method. Finally, the results of the training and test procedure we obtained with our studies with children and teenagers have provided a mass of interesting information to continue improving our knowledge about numerical cognition and mathematical education. This is the content of the current work:

In Study 1 (Chapter 2), we trained 7-8-year-old children with a three-week Approximate Number System training regime based on *Panamath* game, comparing its efficiency with a control group of children exposed to a computer training schedule not involving mathematical activities.

In Study 2 (Chapter 3), we compared the same ANS-based training regime with a novel, digit-quantity-mapping training regime, the *Digits* game, for the same training length and at the same ages.

In Study 3 (Chapter 4), we tested pupils from 8 to 16 years of age. We assessed their abilities at playing both games, extracting information about their ANS accuracy and their Digit-quantity mapping precision, and studied the relation between such abilities and how they correlate with pupil's school marks in mathematics.

1.1.1 Cognitive process in learning mathematics: Numerical Cognition

Thirty years ago, numerical abilities were considered as a derived product of human linguistic competence. In the nineties, however, a disruptive series of scientific papers allowed us to elaborate a different perspective on the cognitive processes involved in numbers. These papers brought to light the existence of a number sense; from them, a novel field of “numerical cognition” was born (Dehaene, 1992).

Since this seminal work, most researchers assume that a preverbal magnitude system, the Approximate Number System (ANS), forms the basis of number processing when we estimate quantities, compare between them, and even allows us to approximate basic additions. We share this system, which has no basis on language, with other animals (Agrillo, Piffer, & Bisazza, 2011; Beran, Evans, & Harris, 2008; Bisazza, Piffer, Serena, & Agrillo, 2010; Dadda, Piffer, Agrillo, & Bisazza, 2009; Dehaene, Dehaene-Lambertz, & Cohen, 1998; Elena, Petrazzini, Agrillo, Izard, & Bisazza, 2016; Jones et al., 2014; Jones & Brannon, 2012; Miletto Petrazzini, Agrillo, Izard, & Bisazza, 2015), people of other cultures (McCrink, Spelke, Dehaene, & Pica, 2013; Pica, Lemer, Izard, & Dehaene, 2004), and preverbal infants (Brannon, 2006; Feigenson et al., 2004; Hyde, 2011; Hyde & Spelke, 2011; Libertus & Brannon, 2009; Xu & Spelke, 2000; Xu et al., 2005; Izard et al., 2008). Even newborn infants seem to have numerical representations (Antell & Keating, 1983; Coubart, Izard, Spelke, Marie, & Streri, 2014; de Hevia, Izard, Coubart, Spelke, & Streri, 2014; Izard et al., 2009).

The number sense is at the very foundation of numerical thinking, because numbers exist to represent magnitudes. However, approximate perceptions of quantities are not a sufficient foundation for our mathematical abilities. Even simple arithmetics calls for a symbolic and accurate mathematical language that allows us to precisely calculate the results of operations, following exact calculation algorithms (Bonny & Lourenco, 2013; Butterworth, 2010; Lemer, Dehaene, Spelke, & Cohen, 2003). Without this exactness, our mathematical competence would remain in the realm of approximation (Dehaene, 2001).

After the discovery of the approximate number system, some studies explored the relation between the ANS and symbolic mathematics. Thus, Dehaene and Cohen (1997) indicated that different levels of numerical skills – notably, approximate skills and exact calculations -- rely on partially separate cerebral circuits. Dehaene and Cohen (1991) presented the case of a patient with a sever lesion in his left hemisphere that impaired his exact calculation abilities, but preserved his approximation abilities. These results, together with others, led Dehaene (1992) to propose the triple-code model of number representation. According to it, numbers are represented at least in three different forms: 1) a visual Arabic code in which numbers are represented as digits, localized in the bilateral inferior ventral occipito-temporal areas; 2) a verbal code in which numbers are represented by words, subserved by the left-hemispheric perisylvian areas, and 3) an analogical quantity or magnitude code, in the bilateral inferior parietal areas (specifically in the intraparietal sulcus IPS where our approximate number system is located) encoding semantical knowledge about numerical quantities. The model included two basics routes by which simple arithmetic problems can be solved. One route is direct and active when the solution to an operation has been acquired by rote verbal learning, such as when we memorize arithmetic facts like multiplications tables or single-digit additions tables. This direct route is blind to the meaning of the numbers manipulated, because only the visual Arabic code (digits) and the verbal code (words) representations are activated, and involves a left cortico-subcortical loop. The second route is the indirect semantic route. In it, the operands of the problem are encoded as quantity representations, acquiring their semantic meaning in the bilateral inferior parietal areas. This second route is used when we must calculate complex additions (including for example 11 + 4), subtractions or divisions. In short, the indirect semantic route is active any time we are solving arithmetic operations that are not normally acquired by rote verbal learning.

These two routes or pathways to solve arithmetic operations (by rote or semantic) use different neural mechanisms that could dissociate as a consequence of cerebral lesions. Indeed, Dehaene & Cohen (1997), presented the case of two patients with different lesions. When damage was in the inferior parietal lobule, the semantical representation of numerical quantities was impaired, affecting subtraction and division more than the single-digit additions or multiplications, especially those with small numbers. This selective loss was consistent with the triple code model because the retrieval of rote verbal knowledge was spared. By contrast, tasks which required the knowledge of the quantities represented by numbers resulted in impaired performance. This was so for numerical comparison, numerical proximity judgements, number bisection tasks, exact calculations (subtraction, division, as well as additions and multiplications with large numbers), or approximations of computations. Thus, the role of the inferior parietal region seems to be critical in performing semantically operations on numerical quantities.

In the second patient, the cerebral lesion was affecting the direct route, the circuit involved in learning by rote arithmetic facts. As a consequence, facts such as multiplication problems were the most impaired. By contrast, tasks involving quantitative number knowledge were spared. This pattern was the opposite to that of the first patient.

Subsequent results suggested that the lateral prefrontal cortex interacts with the intraparietal sulcus in approximate number representations (Nieder & Dehaene, 2009). A recent meta-analyses confirmed that bilateral parietal areas and frontal areas are both involved in number processing (Sokolowski, Fias, Mousa, & Ansari, 2017).

Importantly, the bilateral organization of the parietal areas and the grade of activation of the frontal areas are both affected by development. As for the development of the organization of the parietal areas, in 4 year-old children the involvement of the inferior parietal cortex (and in particular the horizontal Intraparietal sulcus, hIPS) in the representation of numerosity seems to be the same as that in the adult brain (Cantlon, Brannon, Carter, & Pelphrey, 2006). There are several indications that right parietal region may be functional very early in life, before any arithmetic learning, supporting early numerosity abilities (Izard et al., 2008). However, during development there is a progressive shift from the predominance of the right hIPS to a bilateral compensation (Manuela Piazza & Izard, 2009). The left parietal cortex develops with age, acquiring an increased functional specialization in mental arithmetic (Ansari & Dhital, 2006; Rivera, Reiss, Eckert, & Menon, 2005). This may be the result of the development of mathematical language in the left hemisphere, since Arabic digits are more accurately coded than numerosity, in the left parietal area more than in the right hIPS (M. Piazza, Pinel, Le Bihan, & Dehaene, 2007; Manuela Piazza & Izard, 2009). Indeed, in a study with more than 200 adult participants, the only two areas that were found to have increased activity with age were a left inferior/middle temporal region and a left supramarginal/IPS region, related with the representations of numbers (Pinel & Dehaene, 2010). And more recent studies found that in children non-symbolic tasks elicited right IPS activation while more bilateral activation occurred for number words, suggesting that the left IPS plays a greater role in symbolic numerical tasks (Emerson & Cantlon, 2015; Lussier & Cantlon, 2017; Vogel, Goffin, & Ansari, 2015).

Also, the grade of activation of frontal areas changes across development, while the process of learning unfolds. Children show elevated prefrontal cortex activity while processing number symbols. When 6-7 year old children compare numerical symbolic (digits) and nonsymbolic (dots) values, as in adults the occipito-temporal and the parietal cortices were active (Cantlon et al. (2009); however, these tasks also recruited the inferior frontal cortex to a greater degree than in adults. Several studies (e.g. Ansari, Garcia, Lucas, Hamon, & Dhital, 2005) suggest that the frontal cortex participates in forming the initial links between symbolic and nonsymbolic numerical representations, so that children recruit the inferior frontal cortex at an early stage of development to form associations among numerical values at a notation-independent level of abstraction. Developmental changes show a shift from a strong involvement of the right frontal areas towards the intraparietal and posterior parietal regions in processing symbolic magnitudes. This ontogenetic shift in activation may reflect an increasingly flexible mapping between Arabic digits and the quantities they represent. An increasing automaticity of these mappings may require less involvement of the frontal areas, until the parietal regions become fully functionally specialized for processing numerical quantities.

These developmental results are consistent with the progressive automatization and consolidation of the link between Arabic digits and their corresponding quantities from 6 to 22 years of age (Rubinsten, Henik, Berger, & Shahar-Shalev, 2002). While the early stages how the mapping between number words and approximations of sets remains to be fully articulated, there is ample consensus that by the age of 6 children have formed this functional mapping, although the precision and biases involved in this mapping will likely continue to undergo development even into the adult years (Izard & Dehaene, 2008; Sullivan & Barner, 2014).

1.1.2 Two core systems depending on the size of the numerosity

Several studies indicate that our number sense processes estimations differently according to the size of the quantity to be estimated (Feigenson et al., 2004). Our ability to enumerate small numbers of objects, between 1 and 4, is faster and more accurate than when there are more than four objects (Mou & VanMarle, 2014; Trick & Pylyshyn, 1993, 1994). Thus, small numerosities can be rapidly subitized while large numerosities can only be approximately estimated (Dehaene, 1992; Revkin, Piazza, Izard, Cohen, & Dehaene, 2008). These data suggested that different processes may be involved while dealing with small and large quantities (D. C. Hyde & Spelke, 2011; Vuokko, Niemivirta, & Helenius, 2013) These two core systems seems to be present at all ages. What changes across early development is the border between small and large numerosities; going from 2-3 to 3-4 (Coubart et al., 2014). The existence of two different process is also supported by brain imaging research (D. C. Hyde & Spelke, 2009; M. Piazza, Giacomini, Le Bihan, & Dehaene, 2003).

1.1.3 Arabic Digits: naming the digit

Arabic digits are a precise representation of quantities; however, we continue translating them to an approximate representation, accordingly to the accuracy of our approximate number system. This process of representing quantities is automatic (Dehaene, Naccache, et al., 1998; Dehaene & Akhavein, 1995; Dehaene & Naccache, 2001; den Heyer & Briand, 1986; Girelli, Lucangeli, & Butterworth, 2000; Henik & Tzelgov, 1982; Tzelgov, Meyer, & Henik, 1992), irrepressible (Henik & Tzelgov, 1982; Tzelgov et al., 1992), and unconscious (Naccache & Dehaene, 2001a, 2001b; Reynvoet & Ratinckx, 2004). Indeed, whenever we perform tasks involving quantities, it is possible to observe two laws that our ANS follows: the distance effect and the size (or magnitude) effect (Cordes, Gelman, Gallistel, & Whalen, 2001; Dehaene, 2007; Dehaene, Dehaene-Lambertz, et al., 1998; van Oeffelen & Vos, 1982). Now, once the meaning of digits is acquired, the distance effect occurs also when comparing Arabic digits, just as when we directly compare sets of dots. This is evidence that digits are not compared as symbols, but are recoded and computed as quantities (Dehaene, 1992). Thus, both laws, distance and magnitude effects, continue to affect our ability to discriminate digits, whether they be single-digit (Moyer & Landauer, 1967) or two-digit Arabic numbers (Dehaene, Dupoux, & Mehler, 1990; Hinrichs, Yurko, & Hu, 1981; Pinel, Dehaene, Rivière, & LeBihan, 2001).

Recent studies indicated that not only word numerals but also Arabic digits can be named without accessing to its magnitude representation (Herrera & Macizo, 2011). Thus, digits are processed more as words than as pictures, using an asemantic route to access phonological information. It is assumed that the simple reading of number words, like any word, can be done without retrieving their meaning, by a direct connection between the orthography and the phonology of the words (Fias, Reynvoet, & Brysbaert, 2001). On the other hand, naming non-symbolic quantities, like pictures, requires an access to the mental representation of the quantity, that is to say, its meaning (e.g., a dice; Roelofs, 2006). However, an interesting question is the case of numerical symbols such as Arabic digits, which maintain an arbitrary relation both to their phonology and to their meaning. It has been observed access to phonological information when naming Arabic digits without previous access to their meaning (Herrera & Macizo, 2011, 2012). Previous studies in picture naming tasks have shown that if pictures presented belong to the same semantic category, e.g., vehicles (blocked condition), the response times are greater than if pictures presented belong to different semantic categories, e.g., vehicles and animals (mixed condition). This effect has been called semantic interference. However, when the task is to read words, there is no difference between the blocked and the mixed condition (Kroll & Stewart, 1994). This difference between pictures and words or digits is a consequence of the different stages required to access to the phonological information of the word. Therefore, Arabic digits are processed like words, not like pictures, so we can read Arabic digits without activating their meaning. Thus, children may learn to name Arabic digits without this implying an understanding of their meaning.

1.1.4 The number line and the ordering task

Several studies suggest that we represent ordinal sequences spatially organized in an ordered mental line. This mapping occurs both for numerical (Izard & Dehaene, 2008; Odic, Le Corre, & Halberda, 2015; Rouder & Geary, 2014), and non-numerical learned sequences, such as months, letters (Gevers, Reynvoet, & Fias, 2003) or days of the week (Gevers, Reynvoet, & Fias, 2004). The order in which the sequence is mentally represented is activated even when the ordinal position is irrelevant for the task. Furthermore, the fact that a sequence is ordered – even if it is a numerical sequence — does not mean that its ordering maps onto numbers. For example, in the case of the letters, as reported by Gevers et al. (2003), numerically recoding each letter to numbers is unlikely as the translation has a timing cost (Jou & Aldridge, 1999) incompatible with participants responses. Thus, the association between ordinal sequences and space is not necessarily related to numerical values. This is also the case of children who know how to recite numbers. The fact that they know how to order numbers in a line does not mean that they know their numerical meaning. Ordering of the number words, or of the Arabic digits, can be acquired before acquiring the meaning of cardinality. It is a task that can be compared to that of an adult when ordering the letters of the alphabet.

The association of numbers and space may be facilitated by the organization of the brain: numerical and spatial arrangements partially share parietal neuronal networks (Hubbard, Piazza, Pinel, & Dehaene, 2005). Also, there is a strong correlation between spatial and mathematical competences in 12- to 14-years old children (Hermelin & O'Connor, 1986).

The specific direction in which numbers are mentally ordered can be measured by the SNARC effect (Spatial-Numerical Association of Response Codes effect). This shows a preference with faster responses when larger numbers are on the right side or smaller numbers on the left side of the space, indicating a direction of the number mental line that increases from left to right (de Hevia, Vallar, & Girelli, 2008; Dehaene, Bossini, & Giraux, 1993; Hubbard et al., 2005). This direction may respond to a cultural construct. For example when the writing system goes from right to left, the mental representation of digits in space is also reversed, (Dehaene et al., 1993; Zebian, 2005)). Thus, the habits of reading and writing could be the cause of the spatial orientation that we confer to the mental line of numbers. Consistently, in children the SNARC effect has been found to emerge when reading and writing are well consolidated, after the age of 10 (Berch, Foley, Hill, & Ryan, 1999).

It is, however, possible that this bias is much deeper, as left-to-right preferences appear to be present in preverbal infants (Bulf, de Hevia, Gariboldi, & Macchi Cassia, 2017; Rugani & de Hevia, 2017), or even at birth (de Hevia et al., 2014).

This may suggest that some property of the spatial arrangement is directly dependent on low-level responses of the brain.

Several studies indicates that the number mental line undergoes a logarithmic compression, so that large numbers are underrepresented and small numbers overrepresented (Arnaud, Hubbard, Dehaene, & Sackur, 2010; Banks & Hill, 1974; Dehaene, 2007; Shepard, Kilpatric, & Cunningham, 1975). This logarithmic representation of numbers could be proportional to the frequency in which numerals are used, as small numbers are more often read or heard (Dehaene & Mehler, 1992), but not to its cardinality. In the task of placing a number within a segment labeled with 1 at the left and 10 (or 100) at the right, it has been observed that both, the kindergarteners (Berteletti, Lucangeli, Piazza, Dehaene, & Zorzi, 2010; Booth & Siegler, 2006) and the Amazonian Mundurukú (Dehaene, Izard, Spelke, & Pica, 2008), do not place numbers linearly. Small numbers are represented more in the middle, following the logarithmic compression of our number mental line. Improving in this task, that is shifting from a logarithmic to a linear positioning of numbers corresponding more to their actual quantities, occurs latter on development. With education, slowly comes the awareness of the meaning of each number, and of the fact that it is always 1 unit of increment or decrement that separates each number. However, other explanations are possible. Cohen & Sarnecka (2014) used two types of number-lines: the bounded line, that is the traditional line in which both endpoints are labeled (e.g., 0 – 20); and the unbounded line, where only the lower bound (e.g., 0) is marked. They found that developmental changes in performance on the traditionally bounded number line, shifting from logarithmic to linear representation, reflects the improvement in children's measurement skills, rather than changes in their understanding of numerical quantities.

Importantly, a calibration of number mental line can be obtained, not by counting or ordering numbers, but by exposing participants to the information about the quantitative meaning of numbers, such showing how many dots correspond to a given number (Izard & Dehaene, 2008). Arithmetic skills do not seem to benefit from the knowledge of the numerical order per se, but they do seem to improve with the knowledge of the quantity represented by the digits, for example in numerical magnitude comparison tasks (Vogel, Remark, & Ansari, 2014).

Taking all these considerations into account, we propose that ordering numbers is not a reliable way to measure knowledge of cardinality in children. At school, one of the activities that teachers prepare for young children is to order numbers. Success at this task can give the false illusion that children understand what they are doing. More importantly for our purposes, we suggest that the most directly way to improve the comprehension of cardinality, and with it, potentially of mathematical skills, is to create (or reinforce) the mapping between digits and their magnitudes, that is, the quantities they represent.

1.1.5 Acquiring the meaning of numerals and counting

Acquiring the meaning of numerals requires grasping the notion of exact quantities and constructing the mapping between quantities and digits. This process undergoes a long development.

Before counting, but not until two and half years of age, children begin to distinguish numbers from other adjectives, probably in virtue of the syntax of the number words (Wynn, 1990, 1992). Children could then learn the number-word meanings one at a time and in order, possibly by using some form of mutual exclusivity (Sarnecka & Lee, 2009).

For small numerosities, learning to map quantities to their corresponding numerals could not be related to the process of counting since this mapping is prior to the acquisition of the counting principles (Le Corre & Carey, 2007). An important counting principle (there are five principles; Gelman & Gallistel, 1978) is the cardinal principle (CP) that represents the property of the cardinal number of the set: the final item counted in the set represents the number of items in the set. Instead of counting, subitizing seems to be a more primitive tool (Benoit, Lehalle, & Jouen, 2004) since 3 years old children, before they figured out the cardinal principle of counting, are already better on determining the number of objects on short and simultaneous presentation of items than on short but one by one presentations of the same items.

Children are taught to count and, although they know that these numbers are related to quantities, they ignore their exact meaning (Sarnecka & Carey, 2008; Wynn, 1990, 1992). It seems that it takes about six months after the acquisition of counting principles for children to begin mapping numerals beyond “four” to their corresponding quantities. They succeed at this mapping approximately at 5 years of age (Le Corre & Carey, 2007). Often the passage between the acquisition of counting principles and the understanding of how numerals map to quantities is long, extending into a child’s 4^{th} or 5^{th} year of life even for counting words below 20 (Carey, 2004; Wynn, 1990, 1992). Wagner & Johnson (2011) found that children younger than 5 years maintain some correspondence between bigger number words (e.g., “seven” versus “four”) and larger sets of items.

During the process, improvements in counting (that is, in verbal number knowledge) are related to improvements in approximate number system acuity at 3-4 years of age (Shusterman, Slusser, Halberda, & Odic, 2016). It has been suggested that once children master the cardinal principle, they understand that adding one object means to move forward one word in their list of numbers (Sarnecka & Carey, 2008). However, this may not be the case. Le Corre (2014) found that the acquisition of the cardinal principle (CP) does not imply an understanding that the order of number words corresponds to the size of the numerosities they denote. He showed that recent CP-knowers were able to compare number words (e.g., “six” and “ten”) only if they could map them onto non-symbolic quantities. Thus, the activity of counting, by itself, does not imply the understanding of cardinality. Even in special populations, counting without understanding could be associated with a slower mathematical learning process, as in the case for people with Williams syndrome (Ansari et al., 2003; Libertus, Feigenson, Halberda, & Landau, 2014). Similarly, in patient with a lesion in the inferior parietal lobule (where the ANS has been localized), the recitation of ordinal sequences (for example, the alphabet, musical notes, the days of week and months) was perfect. So was numerical series learned by rote like the recitation of the number sequence readily, the even or odd number series until the end of his rote knowledge of this series. However, when the task required jumping of 1, 2, or 3 units back or forth in the number line, a patient with these lesions was severely impaired, and blocks after counting backward for a few items (Dehaene & Cohen, 1997).

In summary, it is important to note that repeating the activity of counting without an understanding of what numerals mean, that is, simply reciting the count list, is not the goal of a numeration system. This ability can even mask the fact that children may lack a real knowledge of numbers and what they mean.

1.1.6 Numerical Cognition and Math performance at School. The studies in the present dissertation.

Two basic cognitive numerical abilities may have an effect on school math performance: The approximate sense of quantities provided by ANS, our non-symbolic and innate sense of number; and the mapping between Arabic digits and the quantities they represent. Accordingly, we centered our research on the relation between these abilities and to math school performance. Our main aim is to see if they could be improved by some form of easy, playful and efficient direct training.

Several studies indicate that there is a relation between ANS and school math performance (Amalric & Dehaene, 2016; Feigenson, Libertus, & Halberda, 2013; Halberda & Feigenson, 2008; Halberda et al., 2012, 2008; Y. He et al., 2016; Libertus, Feigenson, & Halberda, 2013a; Mazzocco, Feigenson, & Halberda, 2011a, 2011b; Shusterman et al., 2016; A. Starr, Libertus, & Brannon, 2013). However, this relation is not always found (Butterworth, 2010; Libertus, Feigenson, & Halberda, 2013b; Sasanguie, Defever, Maertens, & Reynvoet, 2014), though meta-analyses suggest that it is probably a real fact (Chen & Li, 2014). The variability in the results could be explained by the non-linearity of the relation across childhood development (Purpura & Logan, 2015); math anxiety can also be a factor, as in high math-anxious individuals ANS is less precise (M. Isabel Núñez-Peña & Suárez-Pellicioni, 2014); the math profile of the students can also count, given that the correlation between ANS precision and mathematical achievement is stronger for children with a less positive math profile compared to high-performing children (Bonny & Lourenco, 2013).

In our Study 1 (Ferres-Forga, Bonatti, & Halberda, 2017) we performed the first prolonged training of ANS, in a standard school setting. Consistent with the above results, we found that training a non-symbolic comparison task during three weeks at 7-8 year of age does not benefit every child equally, but rather improves the mathematical abilities of those children with a less positive mathematical profile. In addition, the benefits induced by such training are mainly concentrated in symbolic tasks which do not require exact answers, although they do require an implicit understanding of additions, subtractions and multiplications.

In Study 2 (Ferres-Forga, Halberda, & Bonatti, 2017), we took another approach. Following the theoretical motivations presented above, we tried to directly train a sense of cardinality at the same age. We independently trained the digit-to-quantity mapping and the ANS, in order to test whether such training could generate further improvements with respect to ANS training. Our ANS training group repeated the same non-symbolic comparison routine trained in Study 1. It also obtained the same benefits; that is, children improved in symbolic tasks in which an exact answer was not required, although in Study 2 all groups, and not only low-profile children, benefited. In neither studies, did children who underwent ANS training improved their performance in symbolic additions and subtractions when an exact answer was required.

By contrast, in Study 2 the group of participants who were trained in the mapping digits and quantities showed a general improvement in their mathematical competence, whether the tests required exact answers or not. Thus, it seems that strengthening the mapping between digits and quantities generates an improvement on top of any improvement obtained by means of a regime based on training the ANS.

Finally, in Study 3 we test (but not train) 529 students aged between age 8 and 16. Participants performed a one-hour session in which they undertook a quantity discrimination task meant to probe their ANS, and a numerical estimation task assessing their digit-quantity mapping abilities. With this ample sample of participants, we could extend our research in several directions. First, and foremost, we asked which of the two abilities is a better predictive factor of school math performance, as assessed by participants' school math marks.

At the same time, we could verify several other aspects of the relation between math performance and the two systems that our training and tests probed. We could check whether ANS accuracy, as well as the accuracy in mapping Arabic digits to quantities increases with age; we could test whether the relation between the digit-quantity mapping and symbolic math performance that we found in Study 2 at 7-8 years could extended to older ages. Finally, we could have within-participant measures of ANS accuracy and digit-quantity mapping accuracy, a measure we could not collect in our previous studies. Thus, we could also study to what extent these two abilities correlate, and if this correlation depends on the age.

Overall, seven hundred and ten students participated in our studies. Thanks to them, this thesis has been possible.

2. EXPERIMENTAL SECTION 1

2.1 Study 1

Ferres-Forga N, Bonatti L. L, Halberda J. (Submitted*). One-Month of Approximate Number System Training Improves Symbolic Mathematics Performance, But Only in Low-Performing Children.

* This manuscript has been submitted for publication to Journal for Research in Mathematics Education (JRME; Mars 2017).

One-Month of Approximate Number System Training Improves Symbolic Mathematics Performance, But Only in Low-Performing Children

Nuria Ferres-Forga *[1], Luca L. Bonatti [1,2] and Justin Halberda[3]

[1] Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, SPAIN

[2] ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

[3] Department of Psychological and Brain Sciences, Johns Hopkins University, USA

Word count (exc. figures/tables): 10158

*Correspondence should be addressed to Nuria Ferres-Forga (e-mail: nuriaferresforga@gmail.com).

Abstract

We investigated whether training in Approximate Number System precision (ANS) would transfer to improved School Math Performance in 7-9-year-olds compared to a Business as Usual (BAU) Control group. All children participated in Pre- and Post-Training assessments of Addition, Subtraction and Multiplication abilities. During 3-weeks of training (20-minutes per day, two days per week), we found that children in the ANS Training group increased in ANS efficiency both within and across the training days. Individual differences in ANS efficiency were related to math performance. And, ANS training improved math performance beyond BAU for low-performing children. We suggest that in this age range ANS discrimination training may have positive effects on math abilities in low-performing children.

Keywords: approximate number system, early childhood, early mathematics, learning math in children, low-achieving students, math ability, mathematics education, number sense, school mathematics.

1. Introduction

In the classroom, one typically observes a range of mathematical skill from low-performing to high-performing children (Haworth, Kovas, Petrill, & Plomin, 2007; Simzar et al., 2016). Parents, teachers, and community leaders all share a desire to help children improve in school, and perhaps this desire can be most keenly felt for low-performing children who are struggling alongside their middle- and high-performing peers. The mathematical abilities required in the classroom range from simpler skills like intuitive estimation and comparison of numbers of items to more complex abilities such as multiplication. One interest of researchers in cognitive science and education is to understand the range of mathematical content taught in school (e.g., addition, subtraction, multiplication) and the range of basic cognitive abilities that support this content (e.g., the Approximate Number System or Number Sense, Working Memory, Executive Function), and to determine if training in basic cognition can help struggling students improve. This is an area of active research. Here, we tested whether a 3-week computer approximate number activity (twice per week) can improve children's approximate number abilities (in low-middle- and high-performing children) and whether such improvement will transfer to school math skills.

Mathematical skills, even early on, are an important predictor of life outcomes such as health and salary (Parsons & Bynner, 2005; Peters et al., 2014; Ritchie & Bates, 2013). Early mathematical skills also predict later academic achievements in school and in college (Geary et al., 2013; Jordan et al., 2009; Papay et al., 2014; Watts et al., 2014). The importance of the early years in developing mathematical skills has been investigated (Cerda et al., 2015), and results suggest that children who start school behind their peers are at higher risk of being left behind throughout their schooling, resulting in lower academic achievement and continuing difficulties into adulthood (Aubrey, Godfrey, & Dahl, 2006; Geary, 2013; Watts et al., 2014). While improving the performance of students at all levels is important, helping low-performing students may present us with added difficulties because factors such as motivation (Simzar et al., 2016), math anxiety (Pletzer et al., 2015; Z. Wang et al., 2014), lack of a predisposition towards mathematics (Cerda et al., 2015), and the effects of teachers' predictions (Demaray & Elliot, 1998; Tournaki, 2003) come into play. Game-based training of basic math skills may be helpful in this regard as it may be less threatening to children who are low-performing in mathematics.

Classroom-based interventions and teaching have been somewhat less effective in helping low-performing students (Desimone & Long, 2010; Kroesbergen, Luit, & Maas, 2004; Morgan, Farkas, & Maczuga, 2015). However, there is a possibility that even very basic intuitive practice (e.g., with estimating and comparing the numbers of items in collections) could help low-performing children (Bonny & Lourenco, 2013; Bugden & Ansari, 2015; Dyson, Jordan, Beliakoff, & Hassinger-Das, 2015; González et al., 2015). Previous research has found a relationship between individual differences in formal mathematics ability and precision of a basic intuitive sense of number (for review see Feigenson, Libertus, & Halberda, 2013). This relationship can be found at multiple ages: in preschoolers (Bonny & Lourenco, 2013; Gray & Reeve, 2014; Libertus, Feigenson, & Halberda, 2011; Mazzocco et al., 2011b; Moore, vanMarle, & Geary, 2016; Shusterman et al., 2016; A. Starr et al., 2013; Wong, Ho, & Tang, 2016), in primary school children (Gilmore, McCarthy, & Spelke, 2010; Nosworthy, Bugden, Archibald, Evans, & Ansari, 2013; Pinheiro-Chagas et al., 2014), in adolescents (Geary et al., 2013) and in adults (Amalric & Dehaene, 2016; Castronovo & Göbel, 2012). This basic intuitive sense of number comes from a cognitive system called the Approximate Number System (ANS), which supports our estimates of number (e.g., "there are around 100 marbles in that jar", visually estimating), our ordinal comparisons of number (e.g., "this jar has more marbles than that jar"), and basic arithmetic operations over collections (e.g., “I can remove around half of the marbles from this jar to make this jar roughly equal to that jar in number”) (Arrighi, Togoli, & Burr, 2014; Jessica F Cantlon, Platt, & Brannon, 2009). The ANS is present and measurable in young babies (Feigenson et al., 2004; D. C. Hyde, 2011; D. C. Hyde & Spelke, 2011), people of all cultures (McCrink et al., 2013; Pica et al., 2004), all ages (Halberda et al., 2012), and even in other animals (Agrillo et al., 2011; Beran et al., 2008; Bisazza et al., 2010; Dadda et al., 2009; Dehaene, Dehaene-Lambertz, et al., 1998; Jones et al., 2014; Jones & Brannon, 2012). In this sense, the ANS is part of the most basic foundations for our understanding of number and mathematics.

The relationship between the ANS and school math ability appears to hold across the entire life span (Halberda et al., 2012). Preschoolers, students and adults who are able to make more precise and accurate rapid number estimates also tend to do better on tests of formal, written, mathematics (Amalric & Dehaene, 2016; Halberda et al., 2008; Y. He et al., 2016; Libertus et al., 2011). This is true even if one controls for general intelligence, verbal ability, and many other cognitive factors (Libertus et al., 2013a). Early differences in ANS precision also predict later performance in school mathematics (Mazzocco et al., 2011b).

The ANS is not fixed across life. Instead, ANS precision improves throughout the school-age years, attaining its highest precision sometime around age 30 years (Halberda & Feigenson, 2008; Halberda et al., 2012). These developmental improvements highlight a hope that intervening to improve ANS precision might transfer to improvements in school mathematics. Because these basic math intuitions can be built into computer tasks (Halberda et al., 2008; D. C. Hyde, Khanum, & Spelke, 2014; Park & Brannon, 2014), game-based learning that increases ANS precision could help to train these early skills in low-performing students (Bugden et al., 2016; Fuhs & McNeil, 2015; Praet & Desoete, 2014).

Several recent papers suggest that interventions can improve ANS precision and that these improvements may transfer to school mathematics performance. However, the exact nature of the training and the strength of the results are still unknown. For example, a brief (5-minute) intervention to improve ANS confidence transferred to immediate gains in simple arithmetic (J. Wang, Odic, Halberda, & Feigenson, 2016) but the persistence of these gains over an extended period remains unexplored. In other studies, it has also been observed that training in ANS arithmetic (e.g., estimating the results of adding and subtracting clouds of dots) can lead to improvements in symbolic mathematics in children (D. C. Hyde et al., 2014; Obersteiner, Reiss, & Ufer, 2013; Park, Bermudez, Roberts, & Brannon, 2016), in adolescents (Knoll et al., 2016), and in adults (Dewind & Brannon, 2012; Park & Brannon, 2014), but training in simple ordinal comparisons of two collections (e.g., which jar has more marbles) did not show similarly strong transfer results (Lindskog & Winman, 2016; Park & Brannon, 2014, 2016; Pinheiro-Chagas et al., 2014). Thus, it remains to be determined which types of ANS training will be most effective, at which ages, and in which types of children. At the moment, we still don't know which students will benefit most from ANS training. No previous study has focused on the differential effect that ANS training may have on children with different mathematical skills.

In the current study, we focused on second-graders, ages 7- to 9-years old. Formal symbolic school mathematics in second grade includes additions, subtractions, and multiplications, and these operations represent different levels of difficulty for a child; for example additions are easier than subtractions (Knops, Dehaene, Berteletti, & Zorzi, 2014; Linsen, Verschaffel, Reynvoet, & De Smedt, 2014), and subtractions are easier than multiplications (Prado et al., 2011). We wanted to know if three-weeks of ANS training in basic ordinal comparisons (e.g., which jar has more marbles) could enhance mathematical abilities in additions, subtractions, and multiplications and if improvements were related to children's initial mathematical skills.

2. Methods

2.1 Participants

Ninety-one children (44 girls; average age = 7 years 10 months, range from 7 years 3 months to 9 years 3 months) were recruited from Hamelin International Laie School (http://www.hamelininternacionallaie.com/school/). Most children came from families of middle to high socioeconomic status. The study was conducted at the premises of the school. Participants came from four different classrooms. These classes were each served by one of two teachers of mathematics (i.e., each teacher separately taught two classes). There was both a Business as Usual Control group (BAU Control, n= 47) and an Approximate Number System Training group (ANS Training, n= 44), whose difference is explained below. Each teacher was randomly assigned both a BAU Control group and an ANS Training group across their two separate classes. This allowed us to counterbalance treatment and controls across the two teachers. All training activity occurred during the classes' normal computer technology class-time. Computer time and math instruction time were kept separate throughout the study, and teachers and research staff did not highlight or discuss any possible relationship between the computer activity and math class performance with the students.

2.2 Materials

Mathematical competence assessment. In order to determine participants' mathematical competence, we prepared three problem booklets for use as pencil and paper tests. Two of the test booklets contained straightforward arithmetic problems: one *additions* test booklet (Figure 1a) and one *subtractions* test booklet (Figure 1b). For these, the child simply had to write in the correct answer to the addition or subtraction problems. The third test booklet was a novel *operations* test booklet (Figure 1c). In it, each problem was presented with its result, but the operation sign was omitted from the equation. The child had to decide whether the problem solution demanded an addition, a subtraction, or a multiplication sign in order to complete correctly. This third booklet presented a kind of problem that had not been taught before in class.

Because the children's school grades were not detailed enough to allow a classification of children according to their mathematical abilities, we planned to use the *additions* and *subtractions* subtests to classify them into three groups, separating low-, middle- and high-performing children. Then, we used the novel *operations* subtest as a measure of pre- and post-training performance.

Item difficulty, for each of the three subtests, was created under the supervision of the teachers. Two versions of each subtest were prepared, with different problems and different orders, so that we could prepare unique pre- and post-training booklets and counterbalance them across the sessions to control tests effects. Difficulty and problem order were randomized. We created a large number of problems for each subtest. During testing, children were asked to solve as many problems as they could during a 6-minute speeded test. The number of problems in the booklet was such that we could ensure that they could not complete all of them during the allotted time. The *additions* subtest included 210 problems presented on 10 sheets. The *subtractions* subtest included 190 problems presented on 10 sheets presented in a column operation algorithm form (Figure 1a and b). In the *additions* subtest, the maximum number that each addendum could reach was 18, resulting in the highest sum being 18 + 18 and the lowest being 0 + 0. In the *subtractions* subtest, both the minuend and subtrahend ranged between 0 and 18; the result of the subtractions was always positive.

The *operations* subtest included 117 problems presented on 3 sheets (Figure 1c). For each problem, children had to write in the appropriate operation sign. This task included problems of addition, subtraction and multiplication. In the addition problems of the *operations* subtest, the maximum number that the first addendum could reach was 7 and the second 10. In the subtractions problems, both the minuend and subtrahend ranged between 1 and 11, with the result being always positive. In the multiplications problems, only the timetables of 1 to 5 and 10 were used, because those were the timetables children had been exposed to, according to their teachers. The three types of operations (addition, subtraction, multiplication) and the problem difficulties were presented in random order. One of the versions of the *operations* subtest contained forty-two additions, thirty-eight subtractions and thirty-seven multiplications while the other version contained forty-three additions, thirty-nine subtractions and thirty-five multiplications.

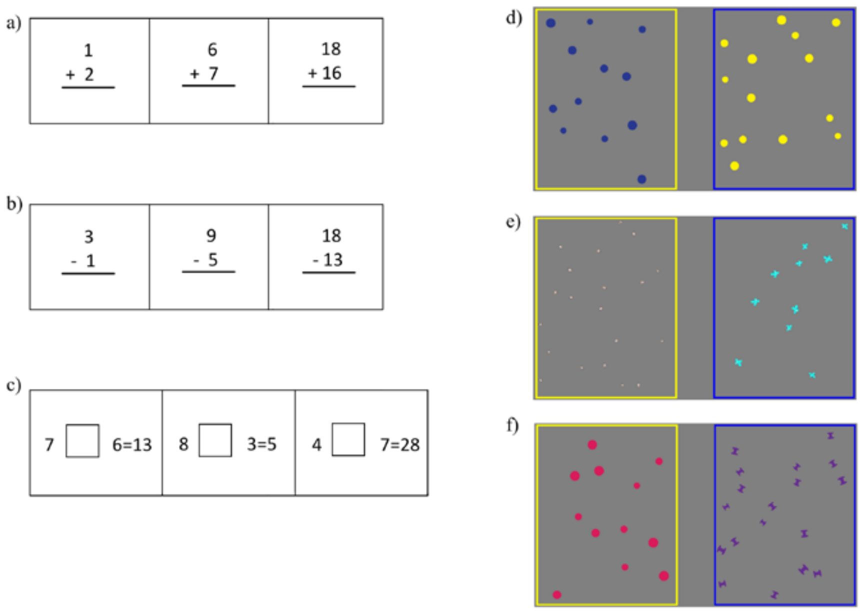


Figure 1. *Examples of problems in additions subtest (a), subtractions subtest (b) and operations subtest (c). Examples of the three models of objects size to control for surface area: size confounded (d), size controlled (e) and stochastic size-control (f).*

Computer Activities. Children trained in the computer classroom of the school, at a fixed schedule (see *Procedure).* In the classroom, twenty-five Hewlett Packard laptops were available (model: HP 620, Pentium (R) Dual Core 2.30GHz, 4GB DRAM, 64-bit; operative system: Windows 7 Home Premium). Each had individual headphones, which the children wore during training. The activities were different for the training and control groups. For the control group, the computer ran two commercial programs, *Tux Paint*, and *Microsoft Word*. For the experimental group, the computer ran a modified version of the computer game Panamath (Halberda et al., 2008), written in Java SE6. In this version, the program would generate trials displaying collections of items contained inside two rectangles appearing on the sides of the screen. For example, twelve teddy bears could appear inside the left rectangle and 6 blue dots inside the right rectangle. The number of items within each rectangle was always between 5 and 21. The items were presented in seven different ratios (larger set/smaller set). The ratios could be 3, 2, 1.5, 1.25, 1.17, 1.14, 1.1. For example, in a 3-ratio trial children could see 21 blue dots vs. 7 yellow dots. Smaller ratios correspond to more difficult trials. On each trial, the items were displayed for 1382 ms.

To vary the relationship between surface area and number, the ANS Training used three different models for controlling object size. Forty-two percent of the trials were size-confounded (or object-size controlled). In them, items average size was equal for both sets, so that cumulative surface area occupied by the objects was congruent with the number of objects (Figure 1d). Forty-two percent of the trails were size-controlled. In them, the average size of the objects was smaller for the larger set, so that the ratio of the cumulative area occupied by the objects in each set was 1 (Figure 1e). In the remaining 16% of trails, object sizes were stochastically varied in order to give children no consistent size cue for number. In these, the average size of the objects varied randomly between being size anti-correlated (where the numerically larger set had less total area on the screen), size-controlled, and size confounded (where the numerically larger set had more total area on the screen). We called these runs stochastic size control (Figure 1f). The training regime of the experimental and control groups is explained in Procedure.

2.3 Procedure

The experiment was run in three phases: *Pre-training*, *Experimental/Control training*, and *Post-training*. The Pre-training and Post-training assessments were intended to measure the mathematical competence of the participants before and after training. These assessments were conducted in the children's respective classrooms, in the presence of their math teacher.

Each of the Pre-training and Post-training assessments required children to answer as many questions as possible in 6 minutes. Children were given 3 booklets containing the addition, subtractions and operation problems. They always began with the additions subtest. After completing this 6-minute test, they had to stop answering and wait until they were given the next test. This second test was always the subtractions subtest, administered with the same 6-minute procedure. The third and final test was the operations subtest. For this test, the teacher had to explain to the children what they were supposed to do more fully, given that the kind of problem presented was new to them. The teacher briefly described the problem structure and completed 3 examples on the board, in front of the class: one example for each of addition, subtraction and multiplication. Children watched as the teacher explained the examples and then they began the 6-minute assessment of the operations subtest.

All tests were administered in Pre-training and Post-training: Pre-training occurred two days before the first training session (initial assessment) and Post-training occurred one day after training completion (final assessment). All tests contained more equations than could be solved in six minutes, so that potential speed and accuracy improvements induced by the training could be assessed by comparing Pre- and Post-test.

The training phase was administered in six different sessions within a three-week period, at a pace of two sessions per week. Both the intervention and control groups were trained on the same days (every Wednesday and Friday), for the same amount of time, and in the same computer classroom. Of the four classes participating in the study, classes A (n=24) and B (n=24) had one of the two math teachers as their yearly teacher, and classes C (n=23) and D (n=20) had the other teacher. Classes B and D composed the control group, while classes A and C composed the intervention group, so that teacher effects could be controlled. Class A was trained on Wednesdays 12 to 12:30 and on Fridays 9:30 to 10; class B on Wednesdays 4:30 to 5pm and Fridays 12 to 12:30; class C on Wednesdays 12:30 to 1pm and Fridays 9 to 9:30; and class D on Wednesdays 4 to 4:30pm and Fridays 12:30 to 1pm.

Control "Business as Usual" (BAU). In the BAU Control group, children practiced *Tux Paint* and *Microsoft Word*. Practicing the latter, they learned to change the type and the size of the fonts, to copy, cut, paste, undo, redo, and add a picture on a document. Practicing the former, they learned the different tools the program offers to draw and modify drawings on screen. Neither activities involved approximate number comparison training. Computer classes started with teachers' instructions, after which children worked individually.

Approximate Number System Training (ANS Training). In the ANS Training group, children practiced the Panamath quantity discrimination game. During training, the computer teacher and the experimenter were always present. Children wore headphones and trained simultaneously. By observation, children appeared to like this game.

The ANS Training presented children with a sequence of pictures, asking them to make ordinal comparisons of the rapidly flashed collections appearing onscreen. On each trial, a child would see two collections of items appearing on the sides of the screen. The child would need to rapidly estimate which side had more items (Figure 1 d-f), typing their answers on a keyboard ("f" and "j" keys for left or right side respectively). Children could not count the items, because the onscreen presentation of the objects was too brief (≈1.3 seconds). Rather, they had to rapidly estimate which of the two sets on each side of the screen had more items. Different collections of items were used on each run of trials, so that the game maintained children's interest. For example, one run could present blue dots vs. yellow dots; another could present cars vs. bears; yet another could display birds vs. dogs. There were 35 trials per run. Children completed approximately 24 runs over the course of three weeks. Feedback was provided after every response: a high-pitched beep indicated a correct answer while a low-pitched beep indicated an incorrect answer.

When introducing the ANS Training game, children were told that they would play a game where they would see some objects (for example, blue and yellow dots) and they would have to choose if there were more blue dots than yellow, or vice-versa, on each trial. They were informed that two different sounds would provide them with feedback about the correctness of their answer. Children were also told that the game would vary in difficulty. They were informed that both speed and accuracy were important in the game. Each participant completed a total of 24 runs except six participants that completed 21 runs, one participant completed 22 runs and one participant completed 20 runs. Each run was comprised of 35 trials (taking approximately 5 minutes of gameplay). Always each run started with the easiest ratio during the first five trials. Then, every five trials the game increased in difficulty, with the ratios becoming closer to 1, until the seven different ratios were presented in the 35 trials forming each run. This procedure was implemented with the aim to increase children's

ANS precision and confidence as observed in previous studies using brief interventions (Odic, Hock, & Halberda, 2012; J. Wang et al., 2016). Children completed the ANS Training during their regular computer class time and it was never described as a math training game. In contrast, the Pre- and Post-Training assessments were completed in the students' regular classroom during regular class time. The teacher and experimenter were careful not to draw any attention or connection between the ANS Training game and the Pre- and Post- paper and pencil assessments.

3. Results

3.1 Approximate number comparison training

First, we asked whether performance on our ANS Training task showed the signatures of engaging the Approximate Number System. The main signature is ratio-dependent performance that results in a specific curve of percent correct as a function of ratio (Feigenson et al., 2004; Libertus & Brannon, 2009; A. B. Starr, Libertus, & Brannon, 2013). Specifically, participants' accuracy at determining the bigger of two approximate numerosities decreases as the ratio between the numbers decreases (i.e. a ratio of 1.1 occurs when the larger set is 11 and the smaller set is 10, and a ratio of 2.0 occurs when the larger set is 20 and the smaller set is 10). This ratio-dependence is predicted by Weber's law and a formal model of discrimination performance predicts a specific curve of percentage of correct answers as a function of ratio. Figure 2 presents of the data from the ANS Training group, separated by the type of size control of the stimuli. The ratio-dependent performance curve is observed for all three of the size-control trial types. That is, as the numerical ratio between the two collections becomes easier (e.g., ratio 3 versus ratio 1.2) children's percentage of correct responses improves. Furthermore, notice that even though overall percent correct is somewhat lower, the curve of ratio-dependent performance is seen for the stochastic size-controlled and size-controlled trials, and children chose the numerically greater collection well above chance.

In all cases, children's performance exhibits the smooth curve of the Approximate Number System. That is, even if size contributes somewhat to children’s decisions, children are making decisions based on number using their ANS. It must be admitted, however, that these three trial types do not exhaust all of the possible ways in which total area, item size, density etc. may contribute to number decisions (Abreu-Mendoza, Soto-Alba, & Arias-Trejo, 2013; Clayton, Gilmore, & Inglis, 2015; L. He, Zhou, Zhou, He, & Chen, 2015; Hollingsworth, Simmons, Coates, & Cross, 1991; Izard & Dehaene, 2008; Leibovich & Henik, 2013; M. Piazza, Izard, Pinel, Le Bihan, & Dehaene, 2004). Such controls were not our main aim here – for our purposes it would be fine if our training game gives children practice on area, items size, density, convex hull and any other of a host of visual cues. As can be seen in the curves of Figure 2 (above 50% performance), children are using these cues to make decisions that agree with set cardinality.

The curves in Figure 2 are generated by fitting a model of Weber’s law to the mean performance of children in each ratio for each size control type. That is, each child contributes equally to the curves, the curves are fit to the group means, and the error bars are ±SE for the group performance.

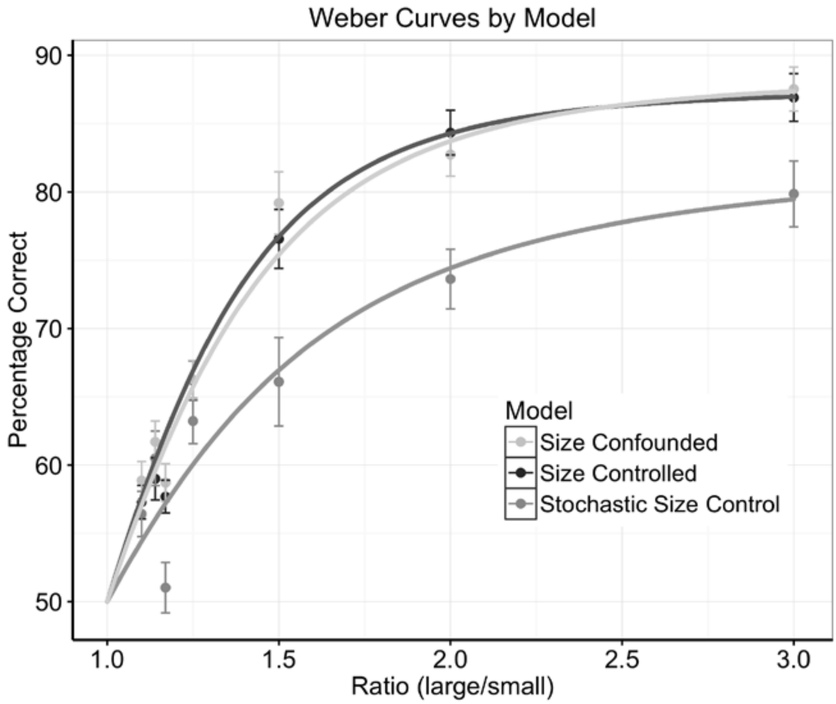


Figure 2. *Accuracy on the ANS acuity task is a function of the ratio between numerical values. X-axis shows the seven different ratios presented during the training (1.1, 1.14, 1.17, 1.25, 1.5, 2, 3). The three ways of manipulating surface-area are shown, each with a corresponding Weber curve, consistent with children making their choice based on number.*

Next, we asked if the ANS Training worked. Did children get better at the ANS Training task across training sessions? Across all ANS training sessions, on average, children responded correctly on 69.24% of trials (SD = 7.85%) and their average RT was 825.22 ms (SD = 340.05 ms). Combining both measures of RT and accuracy, children's performance on the ANS Training task can be analyzed in terms of efficiency, operationalized as the percentage correct divided by the RT. This measure allows us to include both improvements in RT (i.e., getting faster) and improvements in percent correct (i.e., becoming more accurate) in our measure of ANS performance. If children's ANS performance improved across sessions, efficiency should increase during the three weeks of training. Figure 3 presents the measure of efficiency across training sessions. Indeed, children became more efficient at the ANS Training task. The linear fit shows a significant increase in efficiency as children progressed through the ANS training (R^2 = 0.83), showing that the ANS training worked.

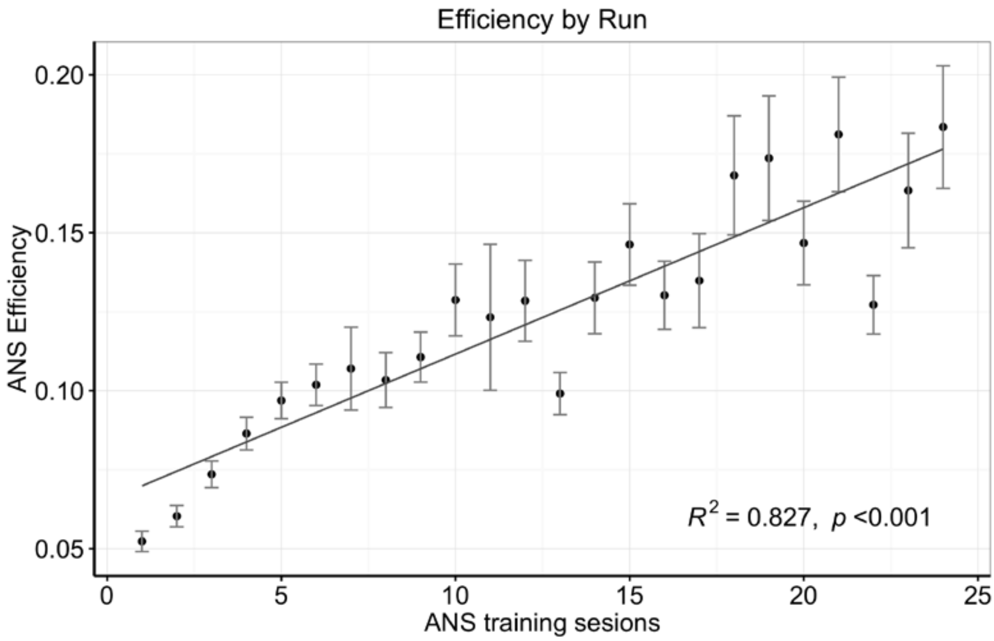


Figure 3. *Efficiency (Percentage correct / RT) on the ANS comparison task as a function of progress in training (composed by 24 runs across three weeks; n=47), The line is a fit of the mean efficiency of the group in each session. Error bars present ±SE for the group performance.*

Next, we asked whether the ANS training also worked on the more local scale of a single day. That is, did children improve within each training day as they progressed through multiple sessions of the ANS Training task? Children always did at least 3 individual runs of the ANS Training task during each day of training. Figure 4 presents the efficiency for the first, second and third run of each day, collapsed across all training days. We ran a one-way between-participant ANOVA with Run (first, second, third) as the factor and Efficiency as dependent variable. We observed a main effect of Run ($F_{(2,138)}$ =5.37, p=0.006). That is, children's efficiency did increase also throughout each day.

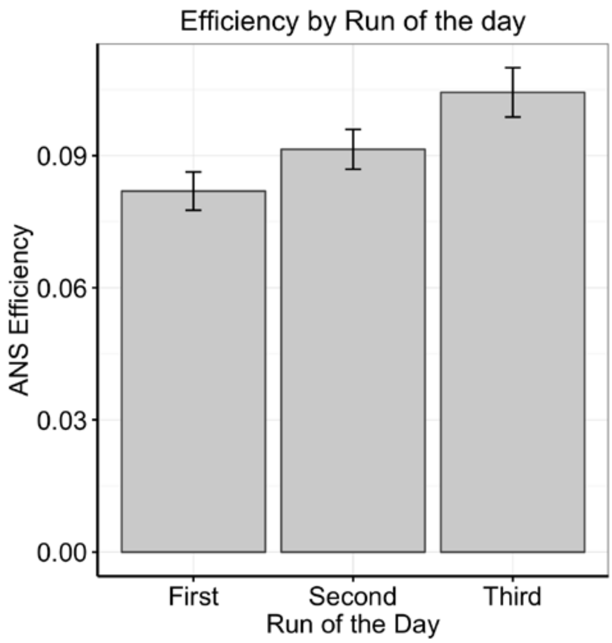


Figure 4. *Efficiency (Percent Correct / RT) during the first, second and third run of each training day, showing improvement in efficiency across runs within each day.*

So far, we have seen that our ANS Training task successfully engaged children in making approximate number decisions (Figure 2), and that the ANS Training task worked to improve children's ANS efficiency both within each day of training (Figure 4) and across the entire 3-weeks of ANS training (Figure 3). Before we ask whether ANS training transferred to symbolic math performance, we want to ask one more important question: is ANS efficiency related to school math performance?

While still a controversial claim, many papers have suggested a relationship between ANS performance and symbolic mathematics performance (e.g., written school mathematics). Because we included subtests for Additions, Subtractions and Operations in our Pre-training measures, we were able to test if performance on these measures of Symbolic Mathematics Performance correlate with children's ANS performance throughout the ANS training. In order to perform this test, we collapsed across all ANS sessions to get one measure of ANS efficiency for each child. We also computed the total number of problems answered correctly across our three Symbolic Mathematics Assessments in Pre-training (Addition, Subtraction, Operations). There is a significant correlation between Symbolic Mathematics Performance in Pre-training and ANS efficiency throughout the ANS Training task ($R^2 = 0.146$, $p = 0.008$; Figure 5). That is, the children with higher symbolic mathematics performance also had higher ANS efficiency. While previous demonstrations of this relationship between ANS acuity and math ability have relied on brief measures of ANS acuity (Booth & Siegler, 2006; Gilmore et al., 2010; Halberda et al., 2012, 2008; Y. He et al., 2016; Libertus et al., 2011; Libertus, Odic, & Halberda, 2012; Lyons & Beilock, 2011; Sasanguie, De Smedt, Defever, & Reynvoet, 2012) here we extend these results to a much longer temporal interval, showing that this relationship holds even when ANS efficiency is measured across three-weeks of training experience.

So far, our results suggest that: 1) our ANS Training task successfully engaged children in making approximate number decisions (Figure 2); 2) the ANS Training task worked to improve children's ANS efficiency both within each day of training (Figure 4) and across the entire 3-weeks of ANS training (Figure 3); and 3) performance on our ANS training task correlated with Symbolic Math Performance (Figure 5). Thus, our design – with Pre- and Post-training measures of Symbolic Math Performance and an effective ANS Training task – will provide a sufficient test of whether ANS training improvements – in basic ordinal comparison – will transfer to symbolic mathematics performance.

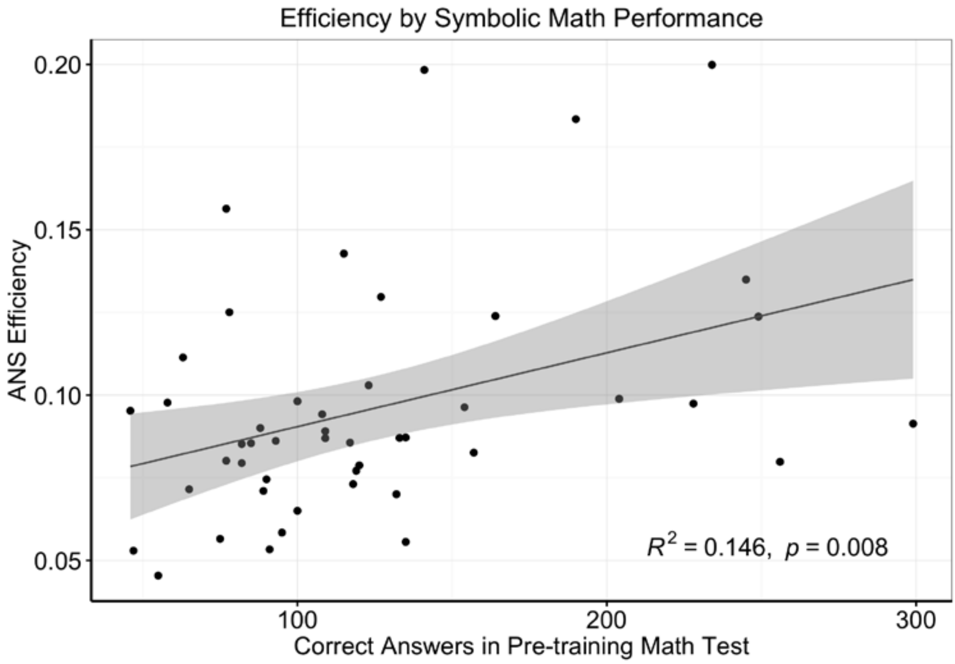


Figure 5. *Correlation between Approximate Number Performance (measured as Efficiency across all sessions of the approximate number comparison task) and Symbolic Math Ability (measured as the number of correct answers in Additions, Subtractions and Operations in Pre-training). Each point represents a child.*

3.2 Pre- and Post-training Tests

Next, we turned to looking at Pre- and Post-training Symbolic Math Performance. First, we considered the Symbolic Math Test (Additions, Subtractions, Operations) collapsed into a single measure of the total number of correct answers across these three subtests for each student. Collapsing across all levels of ability (low- to high-performing students), we found, surprisingly, that training did not have a main effect on Symbolic Math Performance. In a 2 Training Condition (BAU Control, ANS Training) X 2 Time (Pre-, Post-training) ANOVA we found a main effect of Time ($F_{(1,89)}$ =111, p<0.001), no effect of Training Condition ($F_{(1,89)}$ =0.78, p=0.38) and no interaction ($F_{(1,89)}$ =1.76, p=0.19). This suggests that both the BAU Control children and ANS Training children got better at our Symbolic Math Test (perhaps from maturation or retest effects) and that we did not see a global difference between ANS Training and BAU Control children before separating into low-, middle-, and high-performing children (Figure 6).

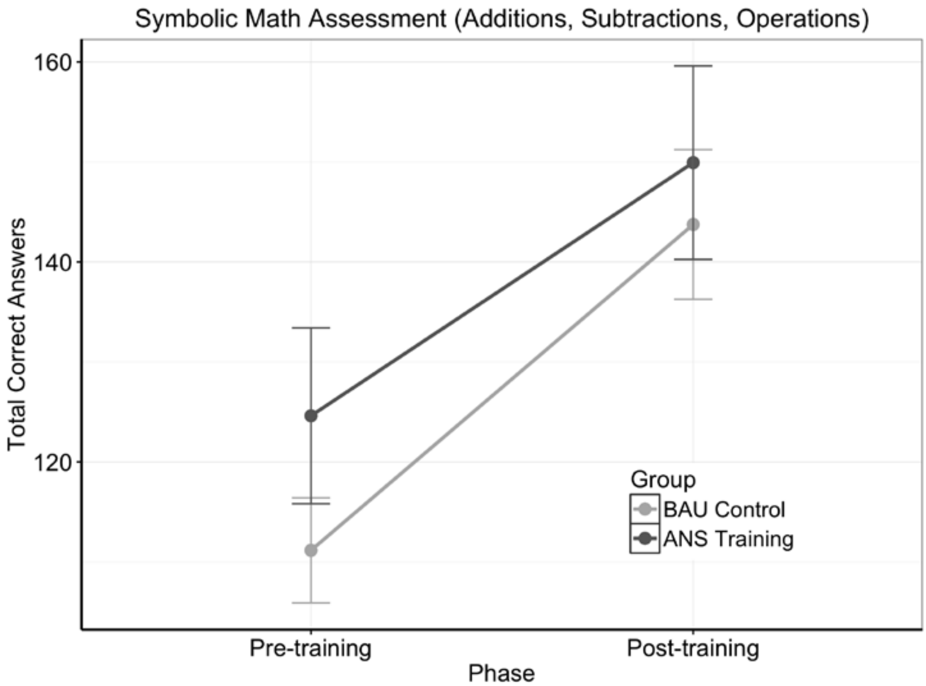


Figure 6. *Total correct answers in Symbolic Math Assessment in Pre- and Post-training, for the BAU Control and ANS Training groups.*

Next, Considering the wide differences between children's mathematical abilities and our interests in the potential differential effect of training on children with different starting points, we investigated the relationship between number of correct answers on the Symbolic Math Test during pre-training (i.e., initial math ability) and percentage growth in symbolic math performance from pre- to post-training (i.e., how much did children improve in symbolic math performance over the course of the month). This relationship can be seen in Figure 7, and a linear regression on these variables returned a significant effect (R^2 = .103, p = .002). Specifically, children who gave fewer correct answers on the pre-training Symbolic Math Test showed higher percentage gains in symbolic math ability from pre- to post-test. Because of these differences in gains across the groups, we next turned to investigating possible heterogeneity in the gain scores for low-, middle- and high-performing children.

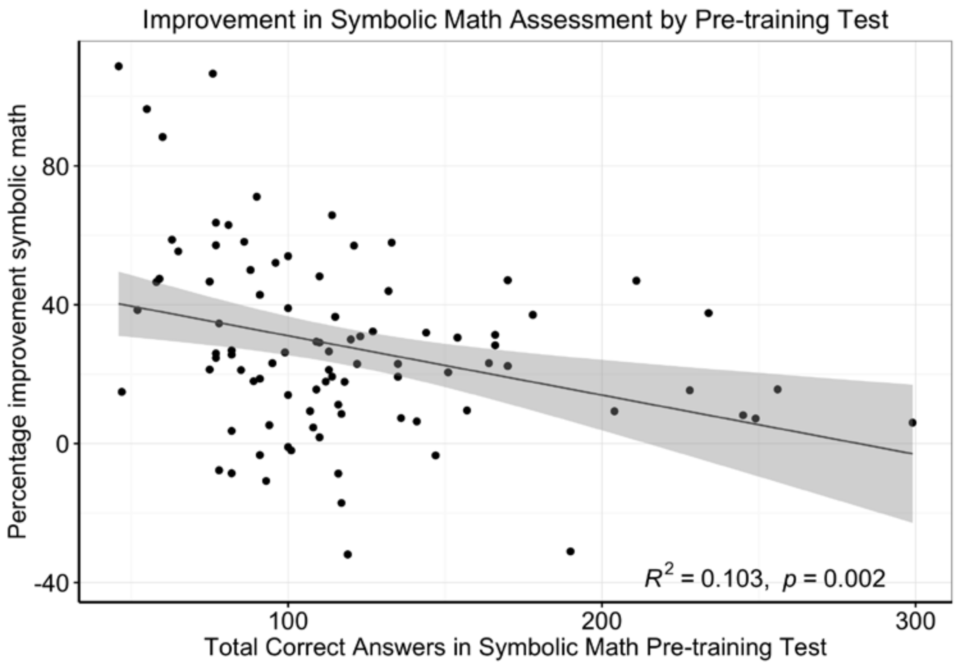


Figure 7. *Percentage improvement in symbolic math for both BAU Control and ANS Training children as a function of Total correct answers in Symbolic Math Assessment in Pre-training.*

To investigate the effect of initial math performance, we split our groups into tertiles according to their results in the pre-training Additions and Subtractions subtests (i.e., the Arithmetic test). We used the Arithmetic test (Additions and Subtractions) to perform a classification into “high”, “middle” and “low” achieving students, because this test is composed of the kinds of problems children are most familiar with and have practiced in school. Indeed, our teachers were currently using students’ addition and subtraction performance in the classroom to evaluate and assign grades. In contrast, the Operations subtest was novel and involved the teacher leading the children through several practice problems during our testing sessions. For this reason, performance on the Operations subtest can be used as an outcome measure, because it assesses how children deal with aspects of mathematical practice that they have not been extensively trained to solve. The Operations subtest also had the nice feature of including addition and subtraction operations as well as multiplications, thereby allowing us to look at the outcomes for each of these types of operations within a novel test.

We grouped all of our children together (N=91) and computed the total number of problems answered correctly on the Symbolic Arithmetic Math Test (Additions and Subtractions subtests) in Pre-training (Figure 8). From this estimate (Figure 8), we grouped children into Low-performing (the lower 33% of all children), Middle-performing (from 33%-66%), and High-performing (the upper 33% of all children). These percentages resulted in the following cutoffs: below 65 correct responses for the Low-performing group, between 66 and 86 for the Middle-performing group, and above 86 correct responses for the High-performing group (Figure 8). This grouping resulted in roughly equal sample sizes in each of our groups of interest ($n_{low\text{-}BAU}$= 15, $n_{low\text{-}ANS}$ = 15, $n_{middle\text{-}BAU}$ = 16, $n_{middle\text{-}ANS}$ = 16, $n_{high\text{-}BAU}$ = 13, $n_{high\text{-}ANS}$ = 16). Because our sample was composed of children attending a non-elite private school in Barcelona, it was representative of the normally-occurring range of mathematical abilities in the Spanish school program, and it did not include supremely gifted students or students who were greatly below the typical level in the tested age class. For this reason, our grouping criteria is likely to capture differences between Low-, Middle- and High-performing students with typical mathematical abilities.

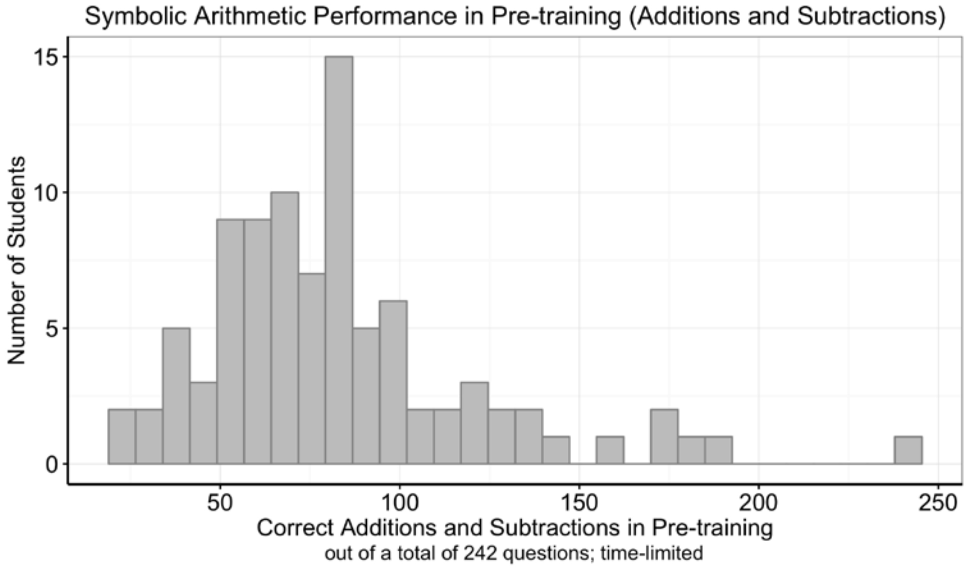


Figure 8. *Histogram for correct answers in the Symbolic Arithmetic pre test (Additions and Subtractions subtests).*

Considering that the grouping variable was measured in the pretest, before any training, we expected to have no significant differences in the average number of correct answers in our Symbolic Math Pre-training Test between groups at any of the three Symbolic Math Levels (Low-, Middle-, High-performing). In a 2 Training Condition (ANS Training, BAU Control) by 3 Symbolic Math Level (Low-, Middle-, High-performing) ANOVA, we found a main effect of Symbolic Math Level ($F_{(2,85)}$ =77.3, p<0.001), which shows that our grouping variable was effective at grouping students by their Pre-training Symbolic Math performance, no effect of Training Condition (p = .5), and a significant interaction of

Training Condition and Symbolic Math Level ($F_{(2,85)}$ =5.53, p=0.005). Bonferroni *post hoc* tests revealed that this interaction was driven by the total correct answers in Math Pre-training being significantly lower for the BAU Control group than for the ANS Training group for High-performing children (p = .03) while other subgroups did not differ (Figure 9). While this appears to be simply a random result of grouping, it is worth bearing in mind that this difference may make it harder to see ANS Training benefits in the High-performing ANS Training group compared to the High-performing BAU Control group.

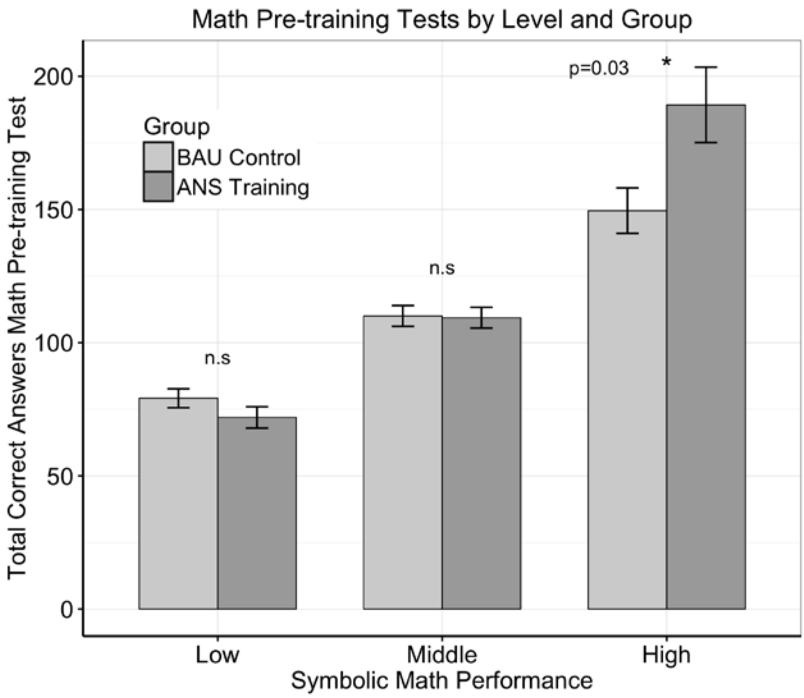


Figure 9. *Total correct answers in Math Pre test for each level (Low, Middle, High) and for each condition (BAU Control and ANS Training). For High-performing children there was a significant difference between BAU Control and ANS Training groups.*

Next, we looked at our outcome variable (the Operations subtest) both before and after training for Low-, Middle-, and High-performing children. Recall that the Operations subtest assessed children's understanding of which operation (addition, subtraction, multiplication) was appropriate within the context of a problem (e.g., Figure 1c). A 2 Training Condition (ANS Training, BAU Control) X 3 Symbolic Math Level (Low-, Middle-, High-performing) X 2 Time (Pre-, Post-training) mixed ANOVA revealed a main effect of Time ($F_{(1,85)}$ =82.14, p<0.001) and of Symbolic Math Level ($F_{(2,85)}$ =22.75, p<0.001) indicating that the three groups of children by math level differed from each other (as expected), and that all children improved with time, regardless of the training regime (Figure 10). Indeed, Bonferroni *post-hoc* tests revealed strong differences from Pre- to Post-training Operations in all groups (Figure 10). However, most importantly, there was a triple interaction between the three factors ($F_{(2,85)}$ =5.74, p=0.004). That is, the training was not equally effective for all groups.

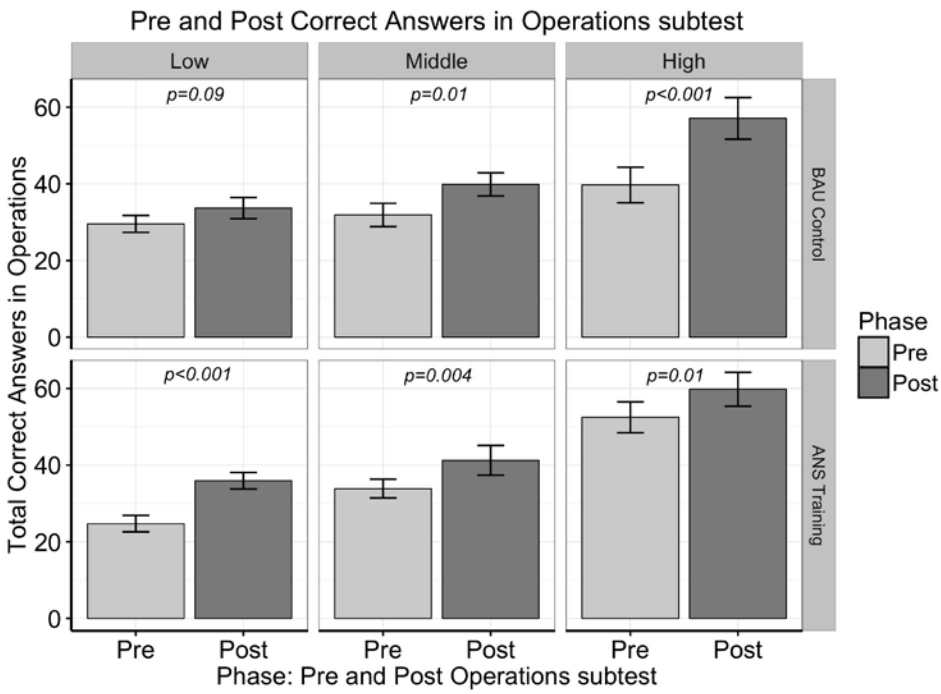


Figure 10. *Total correct answers in Pre and Post Operations subtest by Symbolic Math Levels (Low-, Middle-, High-performing) and Training Condition (BAU Control, ANS Training). (p-values: Bonferroni-corrected).*

In order to better explore this result and more easily compare across groups, we combined Pre- and Post-training scores into percentage change scores (computed as the number of correct responses in Post-training minus the number of correct responses in Pre-training, divided by the number of correct responses in Pre-training). These give a measure of how much each group increased their performance from Pre- to Post-training. One participant provided a percent change in the Operations subtest that exceeded 3 SDs from the mean and we excluded them from this analysis. We ran a two-way between-participants ANOVA with Symbolic Math Level (Low-, Middle-, High-performing) and Training Condition (ANS Training, BAU Control) as factors, and percent change on the Operations subtest as the dependent variable. There were no effects of Symbolic Math Level ($F_{(2,84)}$ =0.859, p=0.4) or Training Condition, ($F_{(1,84)}$ =0.002, p=0.9). However, their interaction was significant ($F_{(2,84)}$ =7.523, p=0.0009; Figure 11). The Bonferroni *post-hoc* tests revealed that the percentage change in the Operations subtest was higher for the ANS Training group than for the BAU Control group in Low-performing children (p= 0.025) (Figure 11), but not for the Middle-performing children (p=0.41). For the Higher-performing children there was an effect in the opposite direction, with the BAU Control group scoring higher than ANS Training group (p=0.004). However, recall, as we already noticed, that for this group differences in both math level were already present in the Pre-training phase (Figure 9), with the BAU Control group scoring lower than the ANS Training group. This was also true for the Operations test, where the High-performing ANS Training children had significantly higher performance compared to BAU Control children ($t(27)$ = 2.09, p = 0.04). The baseline difference on the Operation test between High-performing BAU and

ANS Training children makes it more difficult for the High-performing ANS Training children to show gains in percentage change performance, due to ceiling effects, and may explain why the direction of the training effect was inverse (Figures 10 & 11). No such baseline differences occurred in the Low- and the Middle-performing groups, for which the initial level in the Operation subtest was more balanced. We thus focus on the improvement in the Operations subtest for the ANS Training Low-performing children relative to the BAU Control children.

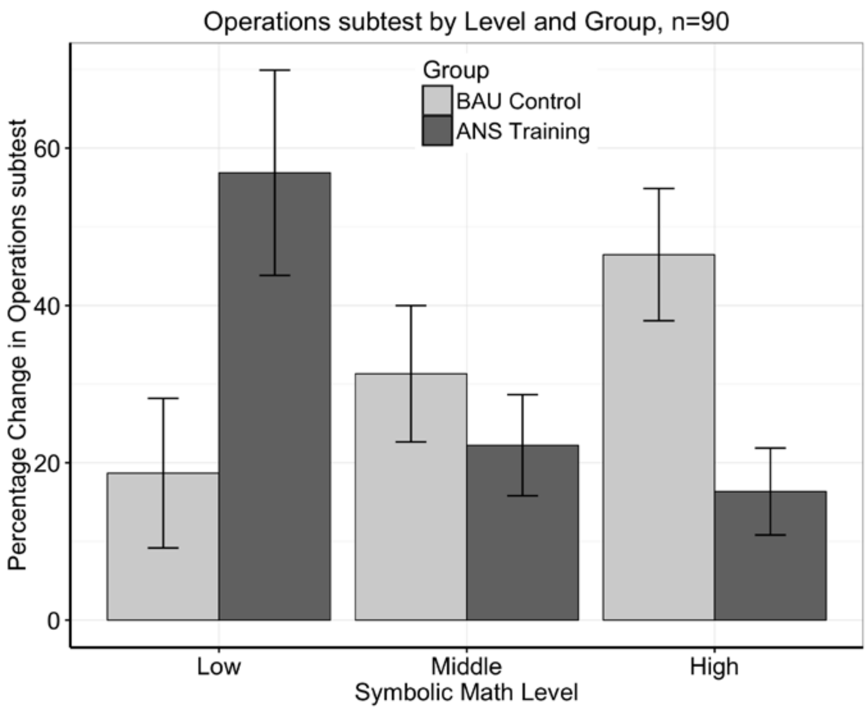


Figure 11. *Percentage change in Operations subtest by Symbolic Math Level (Low-, Middle-, High-performing) and Training Condition (BAU Control, ANS Training).*

For the Low-performing children, we evaluated the percentage change in each of the three problem types of the Operations subtest: Additions, Subtractions, and Multiplications.

We wanted to find out whether Low-performing children would show gains in all three problem types, or in just a subset. We ran a 2 Training Condition (ANS Training, BAU Control) X 3 Problem Type (Additions, Subtractions, Multiplications) mixed ANOVA. It revealed a main effect of Training Condition, with ANS Training children showing greater percentage change than BAU Control children ($p = 0.01$), no effect of Problem Type ($p = 0.91$) and no significant interaction ($p = 0.76$; Figure 12). That is, Low-performing ANS Training children improved on all types of operations in the Operations subtest (Additions, Subtractions, Multiplications), and these improvements exceeded those in the Low-performing BAU Control children. This suggests that ANS Training was successful, beyond Business as Usual (BAU), in training children's ANS and transferring these improvements to Operations performance.

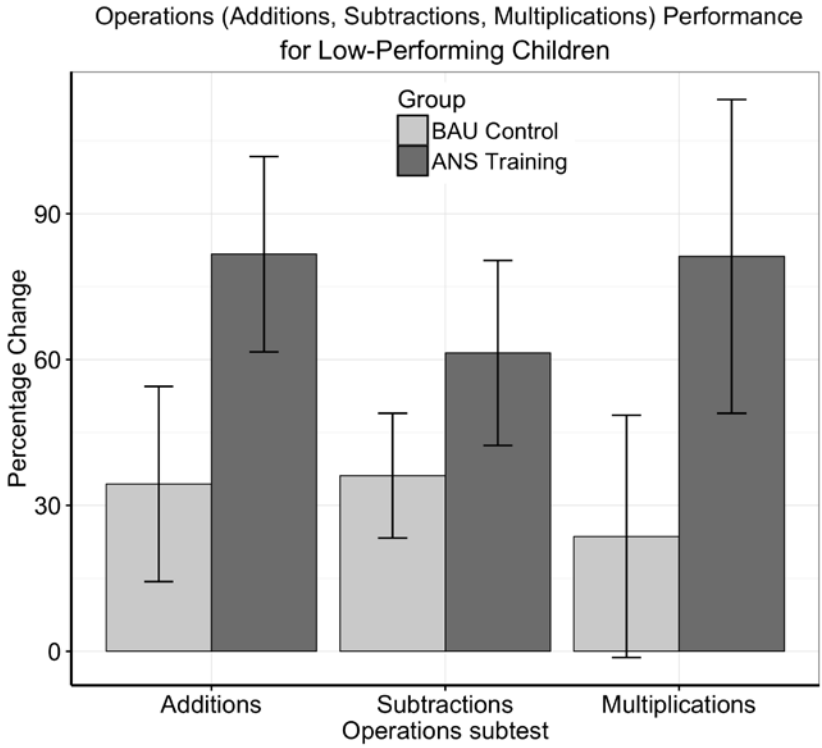


Figure 12. *Percentage change for the three operations (Additions, Subtractions and Multiplications) that appeared within the Operations subtest, for Low-performing Symbolic Math children by Training Condition (BAU Control, ANS Training).*

4. Discussion

There are several aspects of our results that we find very useful in order to understand the effect of ANS training on children of the target age. A first question, which continues to be hotly debated, is whether ANS precision can be effectively measured, and whether individual differences in ANS precision will relate to performance in school mathematics (Clayton & Gilmore, 2014; Patalano, Saltiel, Machlin, & Barth, 2015; Sasanguie et al., 2014). In the present study, we found curves in performance as a function of ratio, even for size-controlled and stochastic size-controlled trials. We firmly believe that continued exploration into the contributions of other dimensions like area, convex hull and density will be invaluable, but this result suggests that our ANS task did measure ANS precision (Figure 2). In the present case, we were interested in simply measuring performance on the ANS task, training this ability, and exploring potential transfers to Symbolic Mathematics Performance. The debate surrounding which visual aspects of a dot display are engaged during a number task will undoubtedly continue. But, notice that if our task happened to train a “convex hull detection ability”, then our results would be an interesting and important demonstration that this training transfers to improved Symbolic Math Performance for Low-performing children.

We found that individual differences in ANS efficiency were related to Symbolic Math performance (Figure 5). We also found that approximately 20-minutes of ANS discrimination practice per day led to significant increases in ANS efficiency over the course of 3-weeks of training (Figure 3) and that efficiency increased across the three sessions within each day (Figure 4). This extends recent results that have shown some trainability of ANS precision (Knoll et al., 2016; Obersteiner et al., 2013; Park et al., 2016; Park & Brannon, 2014).

These successes allowed us to test whether improvements in ANS efficiency would transfer to improved symbolic mathematics performance. Here, our results were more mixed. Overall, both our BAU Control children and ANS Training children improved in Symbolic Math Performance from Pre- to Post-training (Figure 6).

The lack of a specific benefit of ANS training across all levels of math ability could be interpreted as an indication that ANS discrimination training is not particularly relevant for improving Symbolic Math Performance. Two factors lead us to reject this conclusion. First, improvements in math performance could be due to retest improvements, maturation or practice, over and above the specific training we implemented. These independent factors may have obscured any transfer effects we might have detected at the level of the entire group. Because large studies testing ANS training over extended periods of time are not available, it is still premature to assess the relative importance of all these factors.

Second, and most importantly, we did find positive transfer for the most Low-performing children. In the present sample, children who were struggling in symbolic arithmetic relative to their peers showed significantly greater gains in the ANS Training condition compared to the Low-performing children in the BAU Control condition. This motivates us to suggest, quite cautiously, that ANS discrimination training may be more effective for Low-performing children in this age range.

Thus, while supporting and extending some of the recent literature on ANS precision and its relationship with symbolic mathematics, our results are noteworthy for how they may constrain the role of ANS discrimination training as a function of pre-existing math ability. On the one hand, our results may show that ANS discrimination training as an intervention for improving symbolic math performance may not be a general panacea. Other training may be more effective (e.g., non-symbolic arithmetic training; Park & Brannon, 2014). On the other hand, they may indicate that proper training programs require focusing on the particular background knowledge that children possess prior to the training, on the overall school program they attend, or on the particular moment in their development of mathematical knowledge in which the training is implemented. Once we have a clearer understanding of these and other factors, ANS discrimination training may have its proper place in contributing to improvements in mathematical abilities. Future work should continue to determine which types of training work best, adapted to the environmental and institutional conditions, to children's level of knowledge, and to the ages for which such training programs are offered.

Acknowledgments

The research was supported by a grant from the Catalan Government (SGR 2009-1521) to the Grup de Recerca en Neurociència Cognitiva (GRNC) and a McDonnell Scholar Award to J.H.

Special thanks to Sònia Sas, general director of Hamelin Laie International School, for giving access to the participants and the school facilities essential to conduct this research. We thank the teachers Mar Moreno, Tere López and Catherine Teti for their collaboration with the experimenter during the completion of the research.

References

Abreu-Mendoza, R. A., Soto-Alba, E. E., & Arias-Trejo, N. (2013). Area vs. density: influence of visual variables and cardinality knowledge in early number comparison. *Frontiers in Psychology*, *4*, 805. http://doi.org/10.3389/fpsyg.2013.00805

Agrillo, C., Piffer, L., & Bisazza, A. (2011). Number versus continuous quantity in numerosity judgments by fish. *Cognition*, *119*(2), 281–287. http://doi.org/10.1016/j.cognition.2010.10.022

Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences*, 201603205. http://doi.org/10.1073/pnas.1603205113

Arrighi, R., Togoli, I., & Burr, D. C. (2014). A generalized sense of number. *Proceedings. Biological Sciences / The Royal Society*, *281*(1797), 20141791-. http://doi.org/10.1098/rspb.2014.1791

Aubrey, C., Godfrey, R., & Dahl, S. (2006). Early mathematics development and later achievement: Further evidence. *Mathematics Education Research Journal*, *18*(1), 27–46. http://doi.org/10.1007/BF03217428

Beran, M. J., Evans, T. A., & Harris, E. H. (2008). Perception of food amounts by chimpanzees based on the number, size, contour length and visibility of items. *Animal Behaviour*, *75*(5), 1793–1802. http://doi.org/10.1016/j.anbehav.2007.10.035

Bisazza, A., Piffer, L., Serena, G., & Agrillo, C. (2010). Ontogeny of numerical abilities in fish. *PLoS ONE*, *5*(11). http://doi.org/10.1371/journal.pone.0015516

Bonny, J. W., & Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: evidence from the preschool years. *Journal of Experimental Child Psychology*, *114*(3), 375–88. http://doi.org/10.1016/j.jecp.2012.09.015

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*(1), 189–201. http://doi.org/10.1037/0012-1649.41.6.189

Bugden, S., & Ansari, D. (2015). Probing the nature of deficits in the “Approximate Number System” in children with persistent Developmental Dyscalculia. *Developmental Science*. http://doi.org/10.1111/desc.12324

Bugden, S., DeWind, N. K., & Brannon, E. M. (2016). Using cognitive training studies to unravel the mechanisms by which the approximate number system supports symbolic math ability. *Current Opinion in Behavioral Sciences*, *10*, 73–80. http://doi.org/10.1016/j.cobeha.2016.05.002

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, *13*(2), 83–91. http://doi.org/10.1016/j.tics.2008.11.007

Castronovo, J., & Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PloS One*, *7*(4), e33832. http://doi.org/10.1371/journal.pone.0033832

Cerda, G., Pérez, C., Navarro, J. I., Aguilar, M., Casas, J. a, & Aragón, E. (2015). Explanatory model of emotional-cognitive variables in school mathematics performance: a longitudinal study in primary school. *Frontiers in Psychology*, *6*(September), 1363. http://doi.org/10.3389/fpsyg.2015.01363

Clayton, S., & Gilmore, C. (2014). Inhibition in dot comparison tasks. *Zdm*, 1–12. http://doi.org/10.1007/s11858-014-0655-2

Clayton, S., Gilmore, C., & Inglis, M. (2015). Dot comparison stimuli are not all alike: the effect of different visual controls on ANS measurement. *Acta Psychologica*, *161*, 177–84. http://doi.org/10.1016/j.actpsy.2015.09.007

Dadda, M., Piffer, L., Agrillo, C., & Bisazza, A. (2009). Spontaneous number representation in mosquitofish. *Cognition*, *112*(2), 343–348. http://doi.org/10.1016/j.cognition.2009.05.009

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361. http://doi.org/10.1016/S0166-2236(98)01263-6

Demaray, M. K., & Elliot, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, *13*(1), 8–24. http://doi.org/10.1037/h0088969

Desimone, L., & Long, D. (2010). Teacher effects and the achievement gap: Do teacher and teaching quality influence the achievement gap between black and white and high-and low-SES. *Teachers College Record*. Retrieved from http://www.gse.upenn.edu/pdf/desimone/ECLS TCR.pdf

Dewind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: effects of feedback and training. *Frontiers in Human Neuroscience*, *6*, 68. http://doi.org/10.3389/fnhum.2012.00068

Dyson, N., Jordan, N. C., Beliakoff, A., & Hassinger-Das, B. (2015). A Kindergarten Number-Sense Intervention With Contrasting Practice Conditions for Low-Achieving Children. *Journal for Research in Mathematics Education*, *46*(3), 331. http://doi.org/10.5951/jresematheduc.46.3.0331

Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. http://doi.org/10.1016/j.tics.2004.05.002

Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links Between the Intuitive Sense of Number and Formal Mathematics Ability. *Child Development Perspectives*, *7*(2), 74–79. http://doi.org/10.1111/cdep.12019

Fuhs, M. W., & McNeil, N. M. (2015). The Role of Non-Numerical Stimulus Features in Approximate Number System Training in Preschoolers from Low-Income Homes Mary. *Statewide Agricultural Land Use Baseline 2015*, *1*(June), 1–58. http://doi.org/10.1017/CBO9781107415324.004

Geary, D. C. (2013). Early Foundations for Mathematics Learning and Their Relations to Learning Disabilities. *Curr Dir Psychol Sci.*, *22*(1), 43–51. http://doi.org/10.1177/0963721412469398.Early

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' Functional Numeracy Is Predicted by Their School Entry Number System Knowledge. *PLoS ONE*, *8*(1). http://doi.org/10.1371/journal.pone.0054651

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition*, *115*(3), 394–406. http://doi.org/10.1016/j.cognition.2010.02.002

González, M., Kittredge, A., Sánchez, I., Fleischer, B., Spelke, E. S., & Maiche, A. (2015). A way to improve math skills through stimulating ANS. Retrieved December 7, 2015, from http://www.cognicionnumerica.psico.edu.uy/wp-content/uploads/2015/06/articulo_neuroeducacao.pdf

Gray, S. A., & Reeve, R. A. (2014). Preschoolers' dot enumeration abilities are markers of their arithmetic competence. *PLoS ONE*, *9*(4). http://doi.org/10.1371/journal.pone.0094428

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), 1457–65. http://doi.org/10.1037/a0012682

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(28), 11116–20. http://doi.org/10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–8. http://doi.org/10.1038/nature07246

Haworth, C. M. A., Kovas, Y., Petrill, S. A., & Plomin, R. (2007). Developmental origins of low mathematics performance and normal variation in twins from 7 to 9 years. *Twin Research and Human Genetics : The Official Journal of the International*

Society for Twin Studies, *10*(1), 106–117. http://doi.org/10.1375/twin.10.1.106

He, L., Zhou, K., Zhou, T., He, S., & Chen, L. (2015). Topology-defined units in numerosity perception. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41), E5647-55. http://doi.org/10.1073/pnas.1512408112

He, Y., Zhou, X., Shi, D., Song, H., Zhang, H., & Shi, J. (2016). New Evidence on Causal Relationship between Approximate Number System (ANS) Acuity and Arithmetic Ability in Elementary-School Students: A Longitudinal Cross-Lagged Analysis. *Frontiers in Psychology*, *7*(July), 1–8. http://doi.org/10.3389/fpsyg.2016.01052

Hollingsworth, W. H., Simmons, J. P., Coates, T. R., & Cross, H. A. (1991). Perceived numerosity as a function of array number, speed of array development, and density of array Items. *Bulletin of the Psychonomic Society*, *29*(5), 448–450. http://doi.org/10.3758/BF03333967

Hyde, D. C. (2011). Two systems of non-symbolic numerical cognition. *Frontiers in Human Neuroscience*, *5*, 150. http://doi.org/10.3389/fnhum.2011.00150

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, *131*(1), 92–107. http://doi.org/10.1016/j.cognition.2013.12.007

Hyde, D. C., & Spelke, E. S. (2011). Neural signatures of number processing in human infants: evidence for two core systems underlying numerical cognition. *Developmental Science*, *14*(2), 360–71. http://doi.org/10.1111/j.1467-7687.2010.00987.x

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–47. http://doi.org/10.1016/j.cognition.2007.06.004

Jones, S. M., & Brannon, E. M. (2012). Prosimian primates show ratio dependence in spontaneous quantity discriminations. *Frontiers in Psychology*, *3*, 550. http://doi.org/10.3389/fpsyg.2012.00550

Jones, S. M., Pearson, J., DeWind, N. K., Paulsen, D., Tenekedjieva, A.-M., & Brannon, E. M. (2014). Lemurs and macaques show similar numerical sensitivity. *Animal Cognition*, *17*(3), 503–15. http://doi.org/10.1007/s10071-013-0682-3

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*(3), 850–67. http://doi.org/10.1037/a0014939

Knoll, L. J., Fuhrmann, D., Sakhardande, A. L., Stamp, F., Speekenbrink, M., & Blakemore, S.-J. (2016). A Window of Opportunity for Cognitive Training in Adolescence. *Psychological Science*, 956797616671327. http://doi.org/10.1177/0956797616671327

Knops, A., Dehaene, S., Berteletti, I., & Zorzi, M. (2014). Can approximate mental calculation account for operational momentum in addition and subtraction? *Quarterly Journal of Experimental Psychology (2006)*, *67*(8), 1541–56. http://doi.org/10.1080/17470218.2014.890234

Kroesbergen, E., Luit, J. Van, & Maas, C. (2004). Effectiveness of explicit and constructivist mathematics instruction for low-achieving students in the Netherlands. *The Elementary School*. Retrieved from http://www.jstor.org/stable/3202951

Leibovich, T., & Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Frontiers in Psychology*, *4*, 375. http://doi.org/10.3389/fpsyg.2013.00375

Libertus, M. E., & Brannon, E. M. (2009). Behavioral and Neural Basis of Number Sense in Infancy. *Current Directions in Psychological Science*, *18*(6), 346–351. http://doi.org/10.1111/j.1467-8721.2009.01665.x

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, *14*, 1292–300. http://doi.org/10.1111/j.1467-7687.2011.01080.x

Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Is Approximate Number Precision a Stable Predictor of Math Ability? *Learning and Individual Differences*, *25*, 126–133. http://doi.org/10.1016/j.lindif.2013.02.001

Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica*, *141*(3), 373–9. http://doi.org/10.1016/j.actpsy.2012.09.009

Lindskog, M., & Winman, A. (2016). No evidence of learning in non-symbolic numerical tasks - A comment on Park and Brannon (2014). *Cognition*, *150*, 243–247. http://doi.org/10.1016/j.cognition.2016.01.005

Linsen, S., Verschaffel, L., Reynvoet, B., & De Smedt, B. (2014). The association between children's numerical magnitude processing and mental multi-digit subtraction. *Acta Psychologica*, *145*(1), 75–83. http://doi.org/10.1016/j.actpsy.2013.10.008

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*, *121*(2), 256–61. http://doi.org/10.1016/j.cognition.2011.07.009

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, *6*(9), 1–9. http://doi.org/10.1371/journal.pone.0023749

McCrink, K., Spelke, E. S., Dehaene, S., & Pica, P. (2013). Non-symbolic halving in an Amazonian indigene group. *Developmental Science*, *16*(3), 451–62. http://doi.org/10.1111/desc.12037

Moore, A. M., vanMarle, K., & Geary, D. C. (2016). Kindergartners' fluent processing of symbolic numerical magnitude is predicted by their cardinal knowledge and implicit understanding of arithmetic 2years earlier. *Journal of Experimental Child Psychology*, *150*, 31–47. http://doi.org/10.1016/j.jecp.2016.05.003

Morgan, P. L., Farkas, G., & Maczuga, S. (2015). Which Instructional Practices Most Help First Grade Students with and without Mathematics Difficulties? *HHS Public Access*, *37*(2), 184–205. http://doi.org/10.3102/0162373714536608

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PloS One*, *8*(7), e67918. http://doi.org/10.1371/journal.pone.0067918

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction*, *23*(1), 125–135. http://doi.org/10.1016/j.learninstruc.2012.08.004

Odic, D., Hock, H., & Halberda, J. (2012). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology. General*, *142*(1), 255–65. http://doi.org/10.1037/a0030825

Papay, J. P., Murnane, R. J., & Willett, J. B. (2014). High-School Exit Examinations and the Schooling Decisions of Teenagers: Evidence From Regression-Discontinuity Approaches. *Journal of Research on Educational Effectiveness*, *7*(1), 1–27. http://doi.org/10.1080/19345747.2013.819398

Park, J., Bermudez, V., Roberts, R. C., & Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *Journal of Experimental Child*

Psychology, *152*, 278–293. http://doi.org/10.1016/j.jecp.2016.07.011

Park, J., & Brannon, E. M. (2014). Improving arithmetic performance with number sense training: an investigation of underlying mechanism. *Cognition*, *133*(1), 188–200. http://doi.org/10.1016/j.cognition.2014.06.011

Park, J., & Brannon, E. M. (2016). How to interpret cognitive training studies: A reply to Lindskog & Winman. *Cognition*, *150*, 247–251. http://doi.org/10.1016/j.cognition.2016.02.012

Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?* London: National Research and Development Centre for Adult Literacy and Numeracy. Retrieved from http://dera.ioe.ac.uk/22317/1/doc_2979.pdf

Patalano, A. L., Saltiel, J. R., Machlin, L., & Barth, H. (2015). The role of numeracy and approximate number system acuity in predicting value and probability distortion. *Psychonomic Bulletin & Review*, *22*(6), 1820–9. http://doi.org/10.3758/s13423-015-0849-9

Peters, E., Hart, P. S., Tusler, M., & Fraenkel, L. (2014). Numbers matter to informed patient choices: a randomized design across age and numeracy levels. *Medical Decision Making : An International Journal of the Society for Medical Decision Making*, *34*(4), 430–42. http://doi.org/10.1177/0272989X13511705

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555. http://doi.org/10.1016/j.neuron.2004.10.014

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science (New York, N.Y.)*, *306*(5695), 499–503. http://doi.org/10.1126/science.1102085

Pinheiro-Chagas, P., Wood, G., Knops, A., Krinzinger, H., Lonnemann, J., Starling-Alves, I., … Haase, V. G. (2014). In how many ways is the approximate number system associated with exact calculation? *PloS One*, *9*(11), e111155. http://doi.org/10.1371/journal.pone.0111155

Pletzer, B., Kronbichler, M., Nuerk, H. C., & Kerschbaum, H. H. (2015). Mathematics anxiety reduces default mode network deactivation in response to numerical tasks. *Frontiers in Human Neuroscience*, *9*(April), 1–12. http://doi.org/10.3389/fnhum.2015.00202

Prado, J., Mutreja, R., Zhang, H., Mehta, R., Desroches, A. S., Minas, J. E., & Booth, J. R. (2011). Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. *Human Brain Mapping*, *32*(11), 1932–1947. http://doi.org/10.1002/hbm.21159

Praet, M., & Desoete, A. (2014). Enhancing young children's arithmetic skills through non-intensive, computerised kindergarten interventions: A randomised controlled study. *Teaching and Teacher Education*, *39*, 56–65. http://doi.org/10.1016/j.tate.2013.12.003

Ritchie, S. J., & Bates, T. C. (2013). Enduring Links From Childhood Mathematics and Reading Achievement to Adult Socioeconomic Status. *Psychological Science*, *24*(7), 1301–1308. http://doi.org/10.1177/0956797612466268

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology*, *30*(2), 344–357. http://doi.org/10.1111/j.2044-835X.2011.02048.x

Sasanguie, D., Defever, E., Maertens, B., & Reynvoet, B. (2014). The approximate number system is not predictive for symbolic number processing in kindergarteners. *Quarterly Journal of*

Experimental Psychology (2006), 67(2), 271–80. http://doi.org/10.1080/17470218.2013.803581

Shusterman, A., Slusser, E., Halberda, J., & Odic, D. (2016). Acquisition of the cardinal principle coincides with improvement in approximate number system acuity in preschoolers. *PLoS One*, *11*(4), e0153072. http://doi.org/10.1371/journal.pone.0153072

Simzar, R. M., Domina, T., & Tran, C. (2016). Eighth Grade Algebra Course Placement and Student Motivation for Mathematics. *HHS Public Access*, *2*(1), 1–26. http://doi.org/10.1177/2332858415625227

Starr, A. B., Libertus, M. E., & Brannon, E. M. (2013). Infants Show Ratio-dependent Number Discrimination Regardless of Set Size. *Infancy : The Official Journal of the International Society on Infant Studies, 18*(6). http://doi.org/10.1111/infa.12008

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18116–20. http://doi.org/10.1073/pnas.1302751110

Tournaki, N. (2003). Effect of Student Characteristics on Teachers' Predictions of Student Success. *The Journal of Educational Research*, *96*(5), 310–319. http://doi.org/10.1080/00220670309597643

Wang, J., Odic, D., Halberda, J., & Feigenson, L. (2016). Changing the precision of preschoolers' approximate number system representations changes their symbolic math performance. *Journal of Experimental Child Psychology*, *147*, 82–99. http://doi.org/10.1016/j.jecp.2016.03.002

Wang, Z., Hart, S. A., Kovas, Y., Lukowski, S., Soden, B., Thompson, L. A., … Petrill, S. A. (2014). Who is afraid of math? Two sources of genetic variance for mathematical anxiety. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *55*(9), 1056–64. http://doi.org/10.1111/jcpp.12224

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's Past is Prologue: Relations Between Early Mathematics Knowledge and High School Achievement. *Educational Researcher (Washington, D.C. : 1972)*, *43*(7), 352–360. http://doi.org/10.3102/0013189X14553660

Wong, T. T. Y., Ho, C. S. H., & Tang, J. (2016). The relation between ANS and symbolic arithmetic skills: The mediating role of number-numerosity mappings. *Contemporary Educational Psychology*, *46*, 208–217. http://doi.org/10.1016/j.cedpsych.2016.06.003

3. EXPERIMENTAL SECTION 2

3.1 Study 2

Ferres-Forga N, Halberda J, Bonatti L. L, (Submitted*). Improving Mathematics Performance in 7-year-old children: training the Mapping from Quantity to Digit

* This manuscript has been submitted for publication to Cognition (August 2017).

Improving Mathematics Performance in 7-year-old children: training the Mapping from Quantity to Digit

Nuria Ferres-Forga[1]*, Justin Halberda[2] and Luca L. Bonatti[1,3]

[1] Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain

[2] Department of Psychological and Brain Sciences, Johns Hopkins University, USA

[3] ICREA, Pg. Lluís Companys 23, 08010 Barcelona, Spain

Word count (exc. figures/tables): 11618

*Correspondence should be addressed to Nuria Ferres-Forga (e-mail: nuriaferresforga@gmail.com).

Abstract

Arabic digits are a precise representation of quantities, which evoke approximate representations based on the approximate number system (ANS). Studies suggest that both exact and approximate abilities may be crucial for school math performance – ANS training can improve children's math performance and understanding the exact cardinalities indicated by Arabic digits may be a gateway to improving arithmetic skills. Here, we explore whether directly training the relation between Arabic digits and the quantities they represent helps children to become more proficient in mathematics. With a three-week computer-trained regime that can be easily added to the school schedule, we show that strengthening the Digit-Quantity relation improves 7-year-old's competence in additions and subtractions, over and above the improvement obtained by a regime based in training ANS precision alone.

Keywords: approximate number system, children, digit-quantity relation, mathematics education, numerical cognition, numerical symbolic and nonsymbolic representations.

1. Introduction

The construction of mathematical competence is a complex task of which we still lack a complete understanding, although much progress has been made in the last 15 years to unveil the complexity of the mental processes involved. The translation of such scientific understanding into practical ways to improve mathematical skills during the crucial ages when basic mathematical abilities are being acquired still has a long way to go. This is not surprising, considering for instance that many factors may affect this achievement, such as gender (Stoet et al., 2016), motivation (Simzar et al., 2016), socioeconomic status (Thien & Ong, 2015; Verdine et al., 2014), language development (Moll et al., 2015), math anxiety (Pletzer et al., 2015), lack of predisposition to mathematics (Cerda et al., 2015), or the effects of teachers' biases (Demaray & Elliot, 1998; Tournaki, 2003). These are all important factors which contribute to success in school mathematics. Here, we will focus our study on some very basic math skills – the ability to estimate a number of items in a collection, to translate this estimate into an Arabic digits notation (numerical estimation) or to determine which of two collections is greater in number (quantity discrimination). These foundational mechanisms of number representation require coordinating skills that may be important for success in school mathematics.

Numerical thinking relies on a *number sense*: an approximate and non-linguistic ability to estimate quantities, to compare between two of them, and to approximate very basic arithmetic operations, also called the Approximate Number System (ANS). We share this system with other non-linguistic animals, adults from other cultures and preverbal infants (Brannon, 2006; Feigenson et al., 2004; Libertus & Brannon, 2009; Pica et al., 2004; Xu & Spelke, 2000).

However, the existence of a number sense is not sufficient to explain our ability to understand and compute exact operations with numbers and to follow exact calculation algorithms (Butterworth, 2010). Indeed, these and other more advanced mathematical abilities require a symbolic and accurate mathematical language (Bonny & Lourenco, 2013; Dehaene, 2001; Gordon, 2004; Lemer et al., 2003; McCrink et al., 2013; Pica et al., 2004).

Researchers have begun to discover some of the relations between the ANS and symbolic math competence – e.g., the discovery of psychophysical algorithms that describe humans' ability to translate between digits and quantity (Odic, Im, Eisinger, Ly, & Halberda, 2015). But this relation between the ANS and symbolic math competence is not perfect and error-free. For example, it has been documented that humans have a tendency to underestimate the number of items in collections (Hollingsworth et al., 1991; Kemp, 1984; Krueger, 1982; Revkin et al., 2008) and are able to renormalize their estimates based on feedback (Izard & Dehaene, 2008). The development of these findings into concrete training programs to help children has yet to be robustly explored.

More generally, many studies have revealed a connection between the Approximate Number System and school math performance (Amalric & Dehaene, 2016; Feigenson et al., 2013; Halberda & Feigenson, 2008; Halberda et al., 2012, 2008; Y. He et al., 2016; Libertus et al., 2013a; Mazzocco et al., 2011a, 2011b; Shusterman et al., 2016; A. Starr et al., 2013). However, this relation is not always found (Butterworth, 2010; Libertus et al., 2013b), though meta-analyses suggest it is a real result (Chen & Li, 2014). One reason for the diversity of results may be that the relation between the ANS and school mathematics may be affected by several factors: non-linearity across childhood development (Purpura & Logan, 2015); math anxiety fluctuations (ANS may be less precise in high math-anxious individuals; Núñez-Peña & Suárez-Pellicioni, 2014); or skill level (the correlation between ANS precision and mathematical competence is stronger in children with lower mathematical scores than in those with higher scores; Bonny & Lourenco, 2013). Ferres-Forga, Bonatti, & Halberda (2017) found that training in a nonsymbolic comparison task (quantity discrimination) for a three-week period in 7-8 year olds did not benefit every student equally, but rather improved the mathematical abilities of those children with low-starting skills (Ferres-Forga, Bonatti, et al., 2017). In addition, these benefits were mainly visible in symbolic tasks which did not require an exact answer, but required an understanding of the nature of mathematical operations tested (additions, subtractions and multiplications). Training the approximate number system may increase the awareness of what these operations do, and, in Ferres-Forga et al's sample (2017), such training might improve operations understanding of the lowest achieving children, but not exact calculation abilities.

Such training studies may be a valuable way to study the relation between approximate and exact math abilities. If one trains the ANS or school math abilities, do these benefits transfer to other math skills? Here we will focus on training the skill of translating from quantity estimates to Arabic digits (i.e., numerical estimation) and on comparing two quantities (i.e., quantity discrimination). While quantity discrimination has been the focus of previous studies (Ferres-Forga, Bonatti, et al., 2017; D. C. Hyde et al., 2014; Libertus & Brannon, 2010; Park & Starns, 2015; A. Starr et al., 2013; J. Wang et al., 2016), numerical estimation has been less studied as a training intervention (Laski & Siegler, 2007).

Exact arithmetic abilities with precise calculation require a system of symbolic numerals, without which we would not be able to perform exact calculations. The process by which we acquire the meaning of numerals is complex and undergoes a long development, both in the creation of exact concepts of quantities and in the establishment of the mapping between quantities and digits. During this process, children are taught to count but the simple fact that children know how to recite the number words is far from a proof that they understand what such words mean (Sarnecka & Carey, 2008; Wynn, 1990, 1992). Often the passage between the acquisition of counting principles and the understanding of how numerals map to quantities is long, extending into a child's 4th or 5th

year of life even for counting words below 20 (Carey, 2004; Wynn, 1990, 1992).

In contrast, the ability to estimate an approximate number of items is a skill that infants appear to have at birth (Coubart et al., 2014) or shortly after birth (Xu & Spelke, 2000). When shown 8 dots on a screen over and over, infants will become bored; but they will recover some of their interest if the number of items is changed (e.g., 16 dots), even controlling for many of the relevant continuous parameters that may be confounded with number (e.g., total area, density) (Xu et al., 2005). The particular visual cues infants rely on in such studies is still to be determined (Abreu-Mendoza et al., 2013; Clayton et al., 2015; L. He et al., 2015; Hollingsworth et al., 1991; Leibovich & Henik, 2013; M. Piazza et al., 2004), but results suggest that infant abilities are continuous with later abilities to estimate and discriminate number throughout childhood and across the lifespan (Halberda & Feigenson, 2008; Libertus & Brannon, 2009).

The relation between these early abilities to estimate approximate number and children's knowledge of the counting words remains a bit mysterious. Le Corre & Carey (2007), using a dot estimation task and a counting task, found that children did not systematically extend their approximate number knowledge to the numbers in their count list until after age 5 years – significantly after they were proficient counters. But Wagner & Johnson (2011), found that children younger than 5 years maintained some correspondence between bigger number words (e.g., "seven" versus "four") and larger sets of items. Lastly, in an attempt to synthesize these disparate findings, Odic, Le Corre, & Halberda (2015) found that mapping from a collection to a number word may be more difficult for young children than mapping from a number word to a constructed set. While the early stages of making this mapping between number words and approximations of sets remain to be fully articulated, all agree that by the age of 6 years children have formed a functional mapping between them (though, the precision and biases involved in this mapping will likely continue to undergo development even into the adult years (Izard & Dehaene, 2008; Sullivan & Barner, 2014).

Arabic digits are a precise representation of quantities, but their translation into an approximate, nonsymbolic system of representation is automatic (Dehaene, Naccache, et al., 1998; Dehaene & Akhavein, 1995; Dehaene & Naccache, 2001; den Heyer & Briand, 1986; Girelli et al., 2000; Henik & Tzelgov, 1982; Tzelgov et al., 1992), irrepressible (Henik & Tzelgov, 1982; Tzelgov et al., 1992), and unconscious (Naccache & Dehaene, 2001a, 2001b; Reynvoet & Ratinckx, 2004). Furthermore, any task with quantities reveals two phenomena that strongly suggest the signature of an ANS conversion: the distance effect and the magnitude effect (Cordes et al., 2001; Dehaene, 2007; Dehaene, Dehaene-Lambertz, et al., 1998; van Oeffelen & Vos, 1982); both phenomena appear also when we discriminate digits, indicating that when we compare Arabic digits we are transcoding them into a nonsymbolic format, probably an analogic and approximate representation (Dehaene, 1992; Dehaene et al., 1990; Hinrichs et al., 1981; Moyer & Landauer, 1967; Pinel et al., 2001).

Several studies suggest that we represent numerical sequences spatially organized along an ordered mental line (Izard & Dehaene, 2008; Odic, Le Corre, et al., 2015; Rouder & Geary, 2014). However, the signature of a spatial response mapping is not exclusively related to numerical magnitudes. Indeed, any ordinal sequence, even sequences such as months, letters (Gevers et al., 2003) or days of the week (Gevers et al., 2004), are represented as spatially organized in an ordered mental line, even if the ordinal position, or the very fact that quantities are being expressed, is irrelevant for the task.

Thus, while ordinality is applicable to any sequence, cardinality is at the very essence of numbers: an accurate understanding of this concept is needed in order to master mathematical operations. This fact suggests that a simple training of how quantities are spatially placed onto an approximate nonsymbolic number line may prove a necessary, but not sufficient, aspect of constructing a functioning number system. What such a training may not provide is a fostering of the exact nature of the relation between a digit and its associated quantity, the cardinality. That this may be achieved is suggested by the fact that the analogical representations of numerosity can be 'calibrated' by strengthening the mapping between symbols and quantities. Izard & Dehaene (2008) showed that by exposing adult participants to a few trials in which they were told the number of dots on display for one single reference point, their accuracy in the estimations of the numbers of dots on screen improved, even for values which were far from the given reference point. Also, importantly, an overestimated reference point induced global overestimation, while an underestimated point induced global underestimation (Izard & Dehaene, 2008). Laski & Siegler (2007) showed that giving categorical information to calibrate big, medium or small numbers promote linear and accurate estimation in kindergartners. Some data also suggest that the ability to compare two numbers digitally presented, correlates with mathematical abilities even when ANS acuity does not (Guillaume, Gevers, & Content, 2016). Even in 6-8 year olds more accurate mapping between Arabic symbols and nonsymbolic representations of number is related to mathematics achievement (Mundy & Gilmore, 2009), and individual children's acquisition of cardinal principle is related to an improvement in ANS acuity (Shusterman et al., 2016). Such results suggest that if children were led to increase their accuracy in assessment of the exact quantities corresponding to digits, thus better calibrating the quantitative meaning of numbers, an induced overall better understanding of the digit-quantity relation may be generated, potentially percolating in a generalized improvement in their exact calculation abilities. In the present study, we develop a training aimed at strengthening the relation between symbolic and nonsymbolic representations of number. Because our previous work suggests that training the approximate nonsymbolic comparison task for a 3-week period improves mathematical abilities in 7-year old low-performing children relative to a non-mathematical control (Ferres-Forga, Bonatti, et al., 2017), we focused on this age class and assigned our ANS quantity discrimination training, whose effectiveness we previously assessed (Ferres-Forga, Bonatti, et al., 2017), to the current control, while exploring more deeply the possible benefits of numerical estimation training (mapping from quantity to digits) above and beyond the benefits of quantity discrimination training.

2. Methods

2.1. Participants

Ninety-one children (38 girls; average age = 7 yrs 9 mos, range = 6 yrs 4 mos-8 yrs 9 mos) participated in the study. The children mostly came from middle-to-high socioeconomic status families. The study was conducted at the Hamelin International Laie School (http://www.hamelininternacionallaie.com/school/), at the premises of the school. Participants attended four different classrooms, taught by two different teachers. Two classes were randomly assigned to the Quantity Discrimination Training group (n= 46) and the other two classes to the Numerical Estimation Training group (n= 45), with the constraint that assignment was counterbalanced across the two teachers. We explain the differences in training below. All participants completed the full training, except for one child in the Numerical Estimation Training group who completed the initial test but completed neither the whole training nor the final assessment and was excluded from analysis. All training activities were integrated in the normal class schedule, during the computer technology class-time, so that, for children, the training regimes appeared to be standard class activities. In this way, any effect due to potential deviations of the experiments from class routine was minimized. Computer training and math class time were kept separate throughout the study. Teachers and research staff did not mention any relation between the computer activity and math class performance to the students.

2.2. Materials

2.2.1. Mathematical competence assessment.

In order to determine participants' mathematical competence, we administered three pencil and paper tests contained in three different test booklets (Ferres-Forga, Bonatti, et al., 2017): an Addition test (Figure 1a); a Subtraction test (Figure 1b); and an Operation test (Figure 1c). When working with the Addition or Subtraction tests, children had to write the exact answer to the addition or subtraction problems presented. These two tests were entirely composed of problems that children were accustomed to in standard mathematical activity at school. By contrast, the Operation test contained a novel kind of problems that children had not been exposed to before. In them, each equation already contained the result of the computation, but the operation sign itself was omitted. Children had to write the operation sign demanded by the problem (an addition, a subtraction, or a multiplication) to answer correctly.

All the problems in the three test booklets were created under the supervision of the class teachers, who helped in adjusting problem difficulty to an adequate level for the students. Two versions of each test were prepared, with different problems and different orders, but with the same basic set of problems. In this way, different booklets could be used for the pre- and post-training tests, and could be counterbalanced across the sessions, so as to control for tests effects. Difficulty and problem order were randomized. During testing, children were asked to solve as many problems as they could during six minutes for each test. We created a large number of problems for each test, so that the children could not complete all of the problems during the allotted time.

The Addition test included 210 problems presented in a column operation algorithm form, printed on 10 pages. The maximum number that each addendum could reach was 18, with the highest sum being 18 + 18 and the lowest sum being 0 + 0 (Figure 1a).

The Subtraction test included 190 problems presented in a column operation algorithm form, printed on 10 pages. Both the minuend and subtrahend ranged between 0 and 18; the result of the subtractions was always positive (Figure 1b).

The Operation test included 117 problems presented in horizontal format, printed on 3 pages (Figure 1c). The problems were equations whose unknown operation could be an addition, a subtraction or a multiplication sign. In the addition equations, the first addendum had a maximum value of 7 and the second of 10. In the subtraction equations, both the minuend and subtrahend ranged between 1 and 11, with the result being always positive. In the multiplication equations, only the timetables of 1 to 5 and of 10 were used, because those were the timetables children had been exposed to, according to their teachers. The three types of operations (addition, subtraction, multiplication) and the problem difficulties were presented in random order. One of the versions of the operations test contained forty-two additions, thirty-eight subtractions and thirty-seven multiplications while the other version contained forty-three additions, thirty-nine subtractions and thirty-five multiplications.

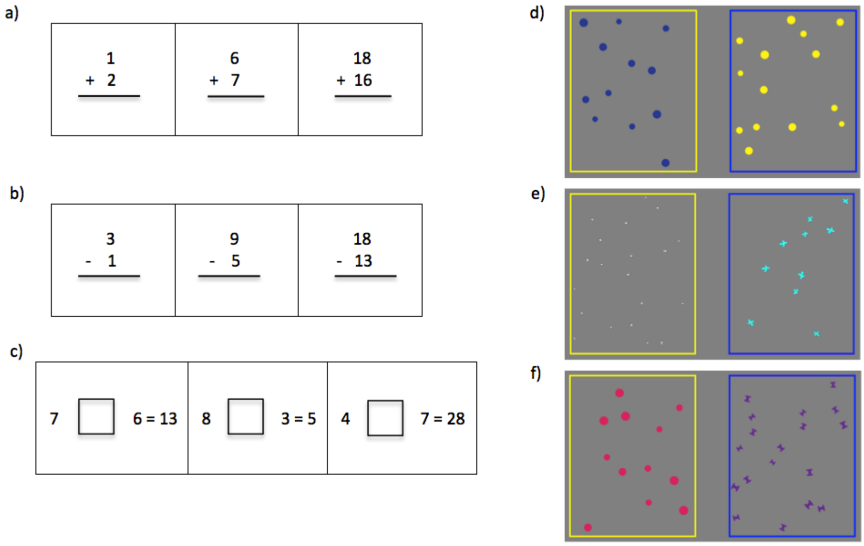


Figure 1. *Examples of problems in additions test (a), subtractions test (b) and operations test (c). Examples of the three models of objects size to control for surface area: size-confounded (d), size-controlled (e) and stochastic size-control (f).*

2.2.2. Computer Activities.

Twenty-eight computers (model: clon PCs, Intel(R) Core(TM) i3-4170 CPU @ 3.70 GHz, 4GB RAM, 64-bit; monitor: 17" LCD 16/9 from ASUS; operative system: Windows 7 Professional) were used for the training activities. The children wore headphones during training. Importantly, while the specific training activities were new to them, children were already acquainted with the material and the training class, so that the training would not appear as an out-of-the ordinary activity. The two experimental groups were trained with different activities.

2.2.2.1. Quantity Discrimination Training.

For this group, the computer ran a modified version of the computer game Panamath (Halberda et al., 2008), written in Java SE6. In this version, the program would generate trials displaying collections of items contained inside two rectangles appearing on the sides of the screen. For example, twelve teddy bears could appear inside the left rectangle and 6 blue dots inside the right rectangle. The number of items within each rectangle was always between 5 and 21. The items were presented in seven different ratios (larger set/smaller set). The ratios could be 3, 2, 1.5, 1.25, 1.17, 1.14, and 1.1. For example, on a 3-ratio trial children could see 21 blue dots in the right side of the screen and 7 yellow dots in the left side. Smaller ratios correspond to more difficult trials. On each trial, the items were displayed for 1382 msec. A run was composed of 35 consecutive trials. To vary the relation between surface area and number, the ANS Training implemented three different models controlling object size: size-confounded (42% of the trials in each run; Figure 1d), size-controlled (42% of the trials; Figure 1e), and stochastic size-control (16% of the trials; Figure 1f). In size-confounded trials, the average size of the items was equal for both sets, so that the cumulative surface area occupied by the objects was congruent with the number of objects. In size-controlled trials, the average size of the objects was smaller for the larger set, so that the ratio of the cumulative area occupied by the objects in each set was equated. In stochastic size-control trials, object sizes were stochastically varied to give children no consistent size cue for number. In these, the average size of the objects varied randomly between size-anti-correlated pictures (in which the numerically larger set occupied less total area on screen), size-controlled pictures, and size confounded pictures (where the numerically larger set occupied more total area on the screen).

2.2.2.1. Numerical Estimation Training.

For this group, the computer ran the "Digits" game, written in PsychoPy v1.83.01. The Digits game program generated two types of trials, the *passive learning trials* and the *active training trials*. A run was composed by 35 consecutive trials; seven of them were passive learning trials and 28 were active training trials. In the passive learning trials, a collection of items was presented on screen while a prerecorded voice named the exact number of items. These trials lasted 1200 ms (Figure 2a). In the active training trials, the program first presented a collection of items for 1 s, in silence (Figure 2a). Then, the collection disappeared and three digits were presented on screen until the child clicked on one of them, ideally the one that represented the quantity of the previous collection of items (Figure 2b).

The purpose of the passive learning trials was to provide children opportunities to directly calibrate their estimation system before the active training trials began (Izard & Dehaene, 2008; Krueger, 1989). Considering that these trials were passive, the program presented them in decreasing order of difficulty (from bigger to smaller sets of items), so that the easier passive learning trials were always presented last, thus increasing children's confidence before the active learning trials started.

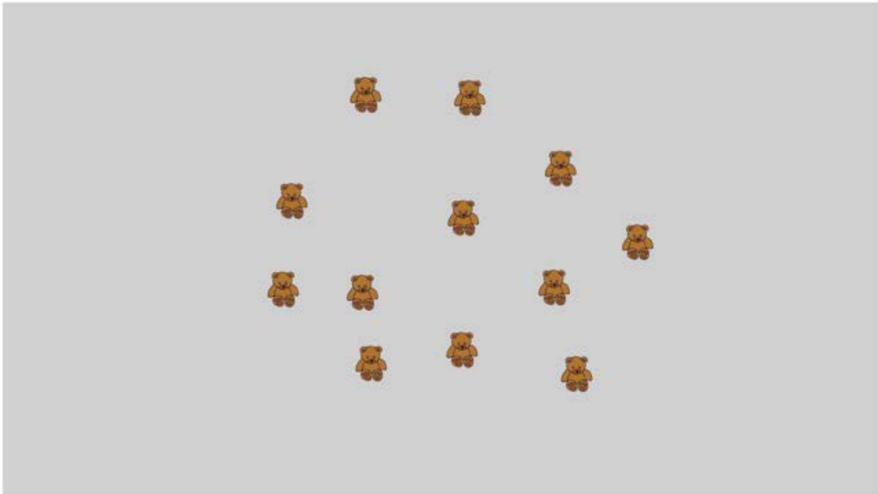
Active training trials were presented in increasing order of difficulty, a procedure which is known to facilitate learning (Odic et al., 2012; J. Wang et al., 2016). Because the distance between digits and the numerosity of the set presented in each trial are subject to known effects (respectively, the distance and size effects (Cordes et al., 2001; Dehaene, 2007; Dehaene, Dehaene-Lambertz, et al., 1998; van Oeffelen & Vos, 1982)) we manipulated these factors in the construction of the game. Thus, the distance between the correct choice (that is, the digit representing the exact number of items) and the distractor choices manipulated the distance. Distances ranged between -6 and 6 from the true value; the greater the distance between the digits, the easier the task of the child. For example, if the target number was 15, a decision between 18, 21, and 15 (or +3, +6, and 0 distance from the correct choice) was easier than the decision between 15, 13, and 17 (0, +2,-2). Thus, throughout each training session, the distances between the correct digit and the other two digits decreased in each run. For the first run, distances could be (0,3,6), (-6,-3,0), or (-3,0,3). We call this set *Span 6*. For the second run, distances could be (0,2,4), (-4,-2,0), or (-2,0,2), or

Span 4. Finally, for the third run distances could be (0,1,2), (-2,-1,0), or (-1,0,1), or *Span 2*.

The size effect was manipulated by incrementing the size of the collections presented in active training trials. Therefore, by manipulating distance and size, the difficulty of the game increased throughout the session and within each run. All set sizes, from 1 item to 21 items, had to be estimated in each of the three spans. Furthermore, the distances for possible answers were maintained irrespective of the target answer, rather than scaling the distractor answers relative to the true answer by some ratio (e.g., larger distances for the competing answers as the target value becomes larger). This was a conscious explicit choice. While scaling the answers is also possible, and might even be preferred in methods used to assess ANS precision rather than training numerical estimation, here we wanted trial difficulty to increase with target number. For our training purposes, this has the goal of providing greater guidance for larger quantities; for example, on a trial with target value 19, the three answer options on a difficult *Span* 2 trial would all provide a strong teaching signal (e.g., 18, 19, 20) for helping to teach children the correct numerical region for an answer (e.g., teaching to overcome an underestimation bias). The position of the correct digit (left, middle, right) as well as its numerical relation with respect to the other two digits (the smallest, in the middle, the largest) were balanced across trials.

In both types of trials (passive learning and active training), the collections of items ranged from 1 to 21, all with the same size and orientation.

a)



b)

Figure 2. *Digits computer game. Example of a learning trial (a) that would be the same for either passive (i.e., with a spoken numerical label "twelve") or active learning trial (i.e., no verbal label), and (b) the response screen for the active learning trial. The collection of items (a) for an active learning trial would appear for 1 s, in silence, then, the collection would disappear and three answer options (b) would be presented on screen until the child clicked on one of them.*

2.3. Procedure

The experiment was run in three phases: *Pre-training Assessment*, *Training*, and *Post-training Assessment*.

2.3.1. Pre-training and Post-training assessments.

These assessments were intended to measure the mathematical competence of the participants before and after training. The Pre-training assessment was administered three days before the first training session. The Post-training assessment was administered three days after training completion. These assessments were conducted in the children's respective classrooms, in the presence of their math teacher. Each of the Pre-training and Post-training assessments required children to answer as many questions as possible in 6 minutes. The assessments always began with the Additions test. After completing this 6-minute test, children had to stop answering, return the additions booklet, and wait until they were given the subtractions booklet. This second, Subtractions test, followed the same 6-minute procedure. The third and final test was the Operations test. For this, the teacher had to explain the task to the children in more detail, given that the kind of problem presented was new to them. The teacher briefly described the equation structures and completed 3 examples on the board, in front of the class: one equation example each for addition, subtraction and multiplication. Children watched as the teacher explained the examples and then they began the 6-minute assessment of the Operations test.

In the Additions and Subtractions Tests, for an answer to be considered correct the child had to write the exact result of the computation. In the Operations Test, for an answer to be correct the child had to write the correct operation sign (additions, subtraction or multiplication).

All tests contained more problems than could be solved in six minutes, so that potential speed and accuracy improvements induced by the training could be assessed by comparing the number of correct answers in Pre- and Post-test.

The Pre- and Post-Training assessments were completed in the students' regular classrooms, during their regular math class time. The teacher and the experimenter carefully avoided drawing any attention to the possible connections between the Quantity Discrimination and Numerical Estimation Training games and the Pre- and Post- paper and pencil assessments.

2.3.2. Training phase.

This phase was administered in six different sessions within a three-week period, at a pace of two sessions per week. Both groups were trained for the same amount of time, in the same computer classroom, during their regular computer class time, and always in the presence of the math teacher of their respective class and the experimenter. After the experimenter gave them the instructions, they worked individually. Children of both groups appeared to like the games included in their training sessions.

2.3.2.1. Quantity Discrimination Training.

In this group children practiced the Panamath quantity discrimination game. This game does not require any understanding of the relation between digits and quantities. On each trial, children saw a picture of two collections of items appear on either side of the screen. They needed to rapidly estimate which side had more items (Figure 1 d-f), typing their answers on a keyboard ("f" and "j" keys for left or right side respectively). Children could not count the items, because the onscreen presentation of the objects was too brief (≈1.3 seconds). Rather, they had to rapidly estimate which of the two sets on each side of the screen had more items. Feedback was provided after every response: a high-pitched beep indicated a correct answer while a low-pitched beep indicated an incorrect answer. Thirty-five consecutive trials formed a run, taking approximately 6-7 minutes each run. Different collections of items were used on each run, so that the game maintained children's interest. For example, one run could present blue dots vs. yellow dots; another could present cars vs. bears; yet another could display birds vs. dogs. Always, the first five trials of each run presented the easiest ratio. Then, every five trials the game increased in difficulty, with the ratios becoming closer to 1 (without ever reaching 1), until the seven different ratios were presented. This procedure was implemented with the aim to increase children's ANS precision and confidence over the course of training, as observed in previous studies using brief interventions (Odic et al., 2012; J. Wang et al., 2016).

When introducing this Discrimination Training game, children were told that they would play a game where they would see some objects -- for example, blue and yellow dots --and would have to choose if there were more blue dots than yellow, or vice-versa. They were informed that two different sounds would provide them with feedback about the correctness of their answer. Children were also told that the game would vary in difficulty. They were informed that both speed and accuracy were important. Each participant completed a total of 24 runs over the course of three weeks, except for six participants who completed 21 runs, one who completed 22 runs and one who completed 20. Considering the minimal amount of training that they lost, these children were kept in the data analysis.

2.3.2.2. Numerical Estimation Training.

To train children's ability to map from approximate to exact quantities when processing digits, children belonging to the Numerical Estimation Training group practiced the Digits game. The training had the same number of runs and trials: 35 trials per run, with approximately 24 runs completed over the course of three weeks. Each run started with seven passive learning trials (Figure 2a). Children needed not take any action during these trials. The following 28 trials composing each run were active training trials (Figures 2ab). The collection of items appearing on screen remained visible for too a short time for children to be able to sequentially count the number of items. Rather, they could respond by giving their best guess. The three digits among which they had to choose remained on screen until the choice was made, by clicking the mouse on one of them. Feedback was provided after every response, with a high-pitch beep for correct answers and a low-pitch beep for incorrect answers, as in the Discrimination Training. Different items were used on each run, in order to keep children interested. And, as explained above, the trial difficulty increased within each run and throughout the session.

When introducing the Numerical Estimation Training game, children were told they would play a game where they would first see a collection of objects for a very short time, while an audio recording would tell them how many objects were in the collection. They were told that they would have to pay attention to these trials, but not take any action. They were also informed that they would then see many trials where a collection would be shown for a short period, after which they would have to decide their numerosity by choosing one of three digits that would appear on screen immediately after the disappearance of the items. They were informed that two different sounds would provide them with feedback about the correctness of their answer. Children were also told that the difficulty of the game would increase within each session. They were informed that both speed and accuracy were important.

3. Results

3.1. Quantity Discrimination Training

The main signature of the Approximate Number System is ratio-dependent performance resulting in a specific curve of percent correct as a function of ratio (Feigenson et al., 2004; Libertus & Brannon, 2009; Starr, Libertus, & Brannon, 2013). Specifically, participants' accuracy at determining the bigger of two approximate numerosities decreases as the ratio between the numbers decreases. This ratio-dependence is predicted by Weber’s law and a formal model of discrimination performance predicts a specific curve of percentage of correct answers as a function of ratio. We checked that performance on our Quantity Discrimination Training task showed this signature. Figure 3 presents the data from the Quantity Discrimination Training group, separated by the type of size control for the stimuli. The ratio-dependent performance curve is observed for all three trial types. That is, as the numerical ratio between the two collections became easier (e.g., ratio 3 versus ratio 1.2) children's percentage of correct responses improved, regardless of the type of size control for the trial. Also, children chose the numerically greater collection well above chance as shown by planned t-tests: Size Confounded, t (45) = 20.34, p < 0.001; Size Controlled, t (45) = 22.76, p < 0.001; Stochastic Size Controlled, t (45) = 12.47, p < 0.001.

In all cases, children's performance exhibits the smooth curve of the Approximate Number System. That is, even if size contributes somewhat to children's decisions, children's numerical decisions were likely based on the ANS. The curves in Figure 3 are generated by fitting a model of Weber's law to the mean performance of children in each ratio for each size control type. That is, each child contributes equally to the curves, the curves are fit to the group means, and the error bars are ±SE for the group performance.

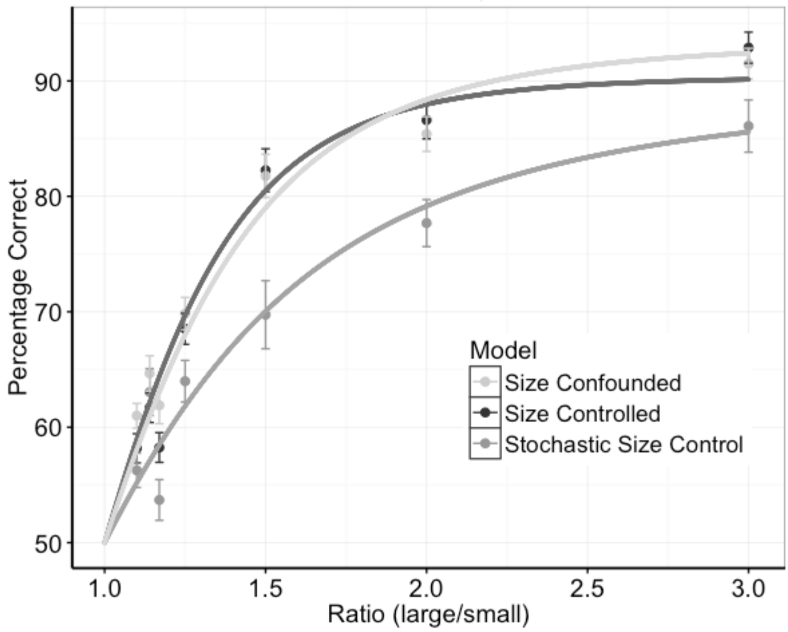


Figure 3. *Accuracy on the Quantity Discrimination Training task as a function of the ratio between numerical values. The x-axis shows the seven different ratios presented during the training (1.1, 1.14, 1.17, 1.25, 1.5, 2, 3). The three ways of manipulating surface-area are shown, each with a corresponding Weber curve, consistent with children making their choice based primarily on number, while being effected somewhat by continuous factors.*

Another measure of performance is children's response time – how long does it take children to make their response on each trial. Best performance in the ANS Discrimination task will occur when participants make the correct choice in as little time as possible. Thus, one indication of improving performance may be decreasing response time across successive runs of the Quantity Discrimination Training task. In Figure 4 we show the mean RT for the group for each run (±SE). A logarithmic training slope was computed for each child. Children had significantly negative training slopes for response time across training sessions: t (45) = -7.3, $p < 0.001$; Figure 4.

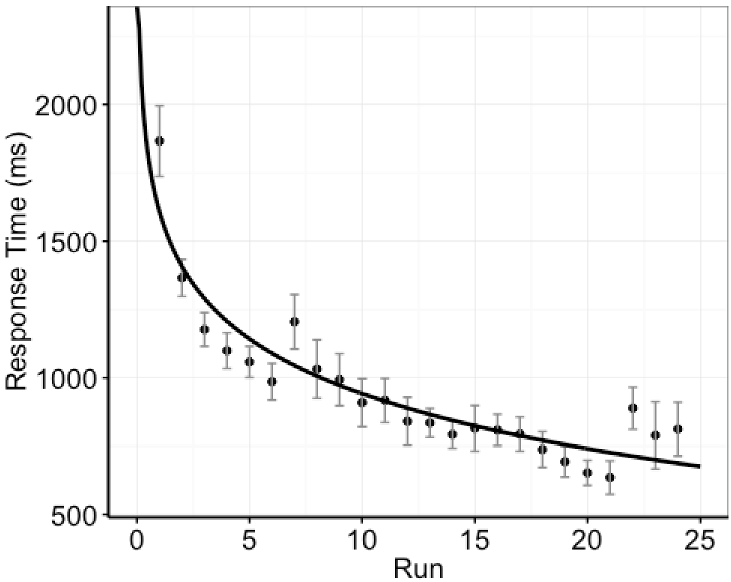


Figure 4. *Response Time (RT) on the Quantity Discrimination Training task as a function of progress in training (composed of 24 runs in six sessions across three weeks). Error bars present ±SE for the group performance. The logarithmic regression line represents the average slope and intercept across participants (n = 46).*

On average, children responded correctly on 72.2% of trials (SD = 7.1%), and their average RT was 952.8 ms (SD = 325.4 ms). Combining both measures of RT and accuracy, children's performance on the Quantity Discrimination Training task can be analyzed in terms of efficiency, operationalized as the percentage correct divided by the RT. This measure allows us to include both improvements in RT (i.e., getting faster) and improvements in percent correct (i.e., becoming more accurate) in our measure of ANS performance. A linear training slope was computed for each child. Children had significantly positive slopes for efficiency across training sessions consistent with their performance improving across training sessions: t (45) = 7.3, $p < 0.001$; Figure 5.

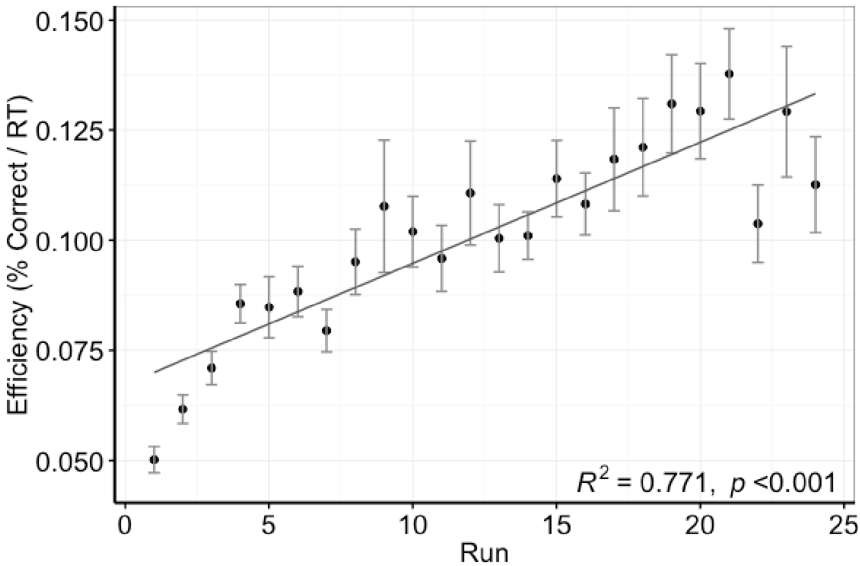


Figure 5. *Efficiency (Percentage correct / RT) on the Quantity Discrimination Training task as a function of progress in training (composed of 24 runs in six sessions across three weeks). Error bars present ±SE for the group performance. The linear regression line represents the average slope and intercept across participants (n = 46).*

3.2. Numerical Estimation Training.

In the Numerical Estimation Training task, children viewed a quantity of briefly flashed items and had to choose the correct match among 3 possible digits. Children completed 3 runs on each day including an Easy (*Span 6*), Medium (*Span 4*) and Hard (*Span 2*) run.

Participants' accuracy across runs is displayed in Figure 6. Consistent with predictions, children did better on the Easy and Medium runs, as seen in a main effect of Difficulty in a 3 Difficulty repeated measures ANOVA: $F_{(2,86)} = 139.9$, $p < 0.001$. Children were also above chance at all levels of difficulty as revealed by planned t-tests: Easy, t (43) = 19.1, $p < 0.001$; Medium, t (43) = 20.1, $p < 0.001$; Hard, t (43) = 12.7, $p < 0.001$.

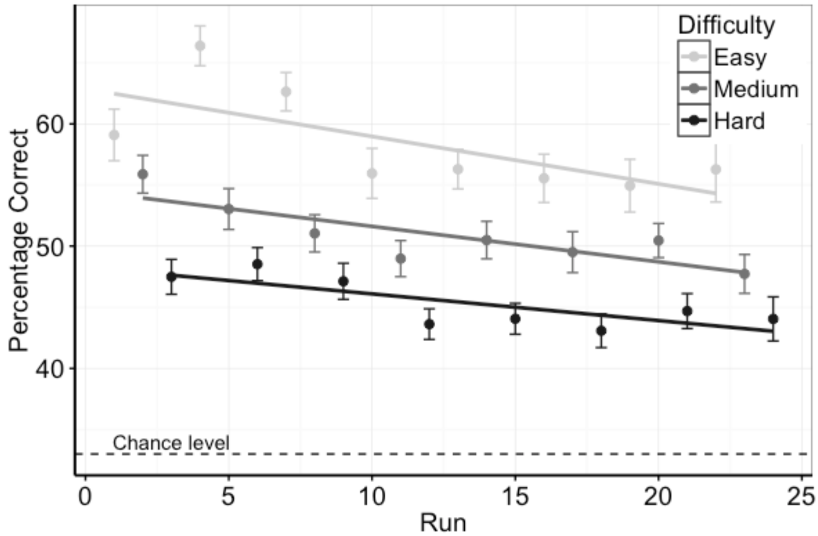


Figure 6. *Percent Correct (chance = 33%) on the Numerical Estimation Training task as a function of progress in training (composed of 24 runs in six sessions across three weeks). Error bars represent ±SE for the group performance. The three linear regression lines (one for each level of Difficulty (Easy, Medium, Hard) represent the average slope and intercept across participants (n = 44).*

Next, we analyzed children's response time – how long it takes children to make their choice on each trial. In Figure 7 we show the mean RT for the group for each run (±SE). A logarithmic training slope was computed for each child. Children had significantly negative training slopes for response time across training sessions: *t* (43) = -3.38, *p* = 0.001; Figure 7.

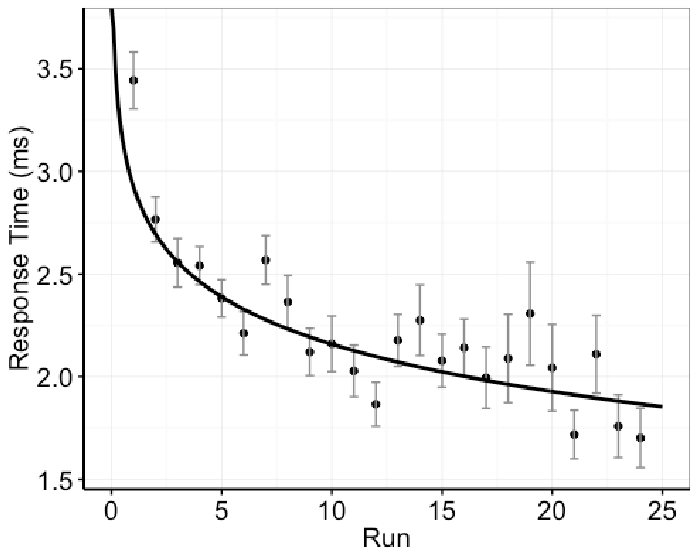


Figure 7. *Response Time (RT) on the Numerical Estimation Training task as a function of progress in training (composed of 24 runs in six sessions across three weeks). Error bars represent ±SE for the group performance. The logarithmic regression line represents the average slope and intercept across participants (n = 44).*

On average, children responded correctly on 51.3% of trials (chance = 33%; SD = 6.3%). Average RT was 2234.6 ms (SD = 615 ms). The efficiency, operationalized as the percentage correct divided by RT, increased as children progressed throughout the sessions of the Numerical Estimation Training (Figure 8). A linear training slope was computed for each child. Children had significantly positive slopes for efficiency across training sessions consistent with their performance improving: t (43) = 2.75, p = 0.008; Figure 8.

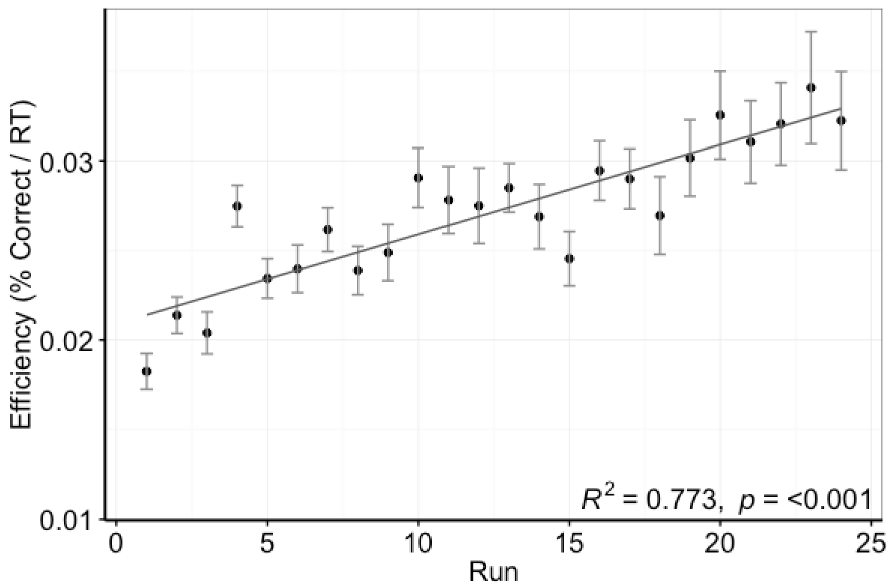


Figure 8. *Efficiency (Percentage correct / RT) on the Numerical Estimation Training task as a function of training runs. Error bars represent ±SE for the group performance at each run. The linear regression line represents the average slope and intercept across participants (n = 44).*

In sum, both the Quantity Discrimination Training and the Numerical Estimation Training showed signatures of engaging the ANS and had the appropriate level of difficulty so that children's task efficiency could improve during the 3 weeks of training.

3.3. Effects of training on symbolic mathematics

The measure of main interest for this study is the effect of training on children's mathematical performance, Pre versus Post-training in Additions, Subtractions and Operations tests.

3.3.1. Additions Test.

A 2 x 2 mixed ANOVA, with Training Condition (Quantity Discrimination or Numerical Estimation Training) as a between-participant factor and Phase (Pre-, Post-training) as a within-participant factor, and total correct answers as the dependent variable, revealed a main effect of phase ($F_{(1,88)}$ =10.03, p=0.002, η^2=0.01) and a significant Training Condition-Phase interaction ($F_{(1,88)}$ =7.03, p=0.009, η^2=0.008). Bonferroni *post hoc* tests revealed that this interaction was driven by the fact that in the Numerical Estimation Training group the total correct answers in the Post-training phase were higher than in the Pre-training phase (respectively, M_{Post} = 43.34, SD = 3.33 and M_{Pre} = 35.6, SD = 3.15, p < .0001; Figure 9), while the improvements for the Quantity Discrimination Training group were not significant (respectively, M_{Post} = 37.74, SD = 3.52 and M_{Pre} = 37, SD = 2.68, p > .05; Figure 9). This suggests that Numerical Estimation Training significantly improved speeded Additions performance while Quantity Discrimination Training did not.

3.3.2. Subtractions Test.

An identical ANOVA was run for correct answers in the Subtractions Test. It revealed a main effect of phase ($F_{(1,88)}$ =34.32, p<0.001, η^2=0.03), a strong tendency towards an effect of training condition ($F_{(1,88)}$ =3.72, p=0.057, η^2=0.037), and a significant Training Condition-Phase interaction ($F_{(1,88)}$ =6.58, p=0.012, η^2=0.006). With respect to the significant interaction, Bonferroni *post hoc* tests revealed that in both training conditions the total correct answers increased from Pre-training to Post-training. Nevertheless, in the Numerical Estimation Training group the total correct answers increased more than in the Quantity Discrimination Training group (Numerical Estimation group, M_{Post} = 47.81, SD = 3.86, M_{Pre} = 37.22, SD = 3.18, p < .0001; Quantity Discrimination group, M_{Post} = 35.98, SD = 3.26, M_{Pre} = 31.69, SD = 2.54, p < .027). In both cases the changes were significant, but they were smaller for the Quantity Discrimination group (Figure 9). And, while there was no difference in total correct answers between the groups in Pre-training (p > .05, Figure 9), children in the Numerical Estimation Training group scored significantly higher in the Post-training (p=0.021; Figure 9).

Thus, both Additions Test and Subtractions Test performance suggest that the Numerical Estimation Training was more effective than the Quantity Discrimination Training for generating significant improvements in children's school math performance after training.

3.3.3. Operations Test.

An identical ANOVA was run for correct answers in the Operations Test. It revealed a main effect of phase ($F_{(1,88)}$ =29.33, p<0.001, η^2=0.04), no effect of group and no interaction (Figure 9). This suggests that both Quantity Discrimination Training and Numerical Estimation Training were effective for improving children's performance on the novel test of Operations.

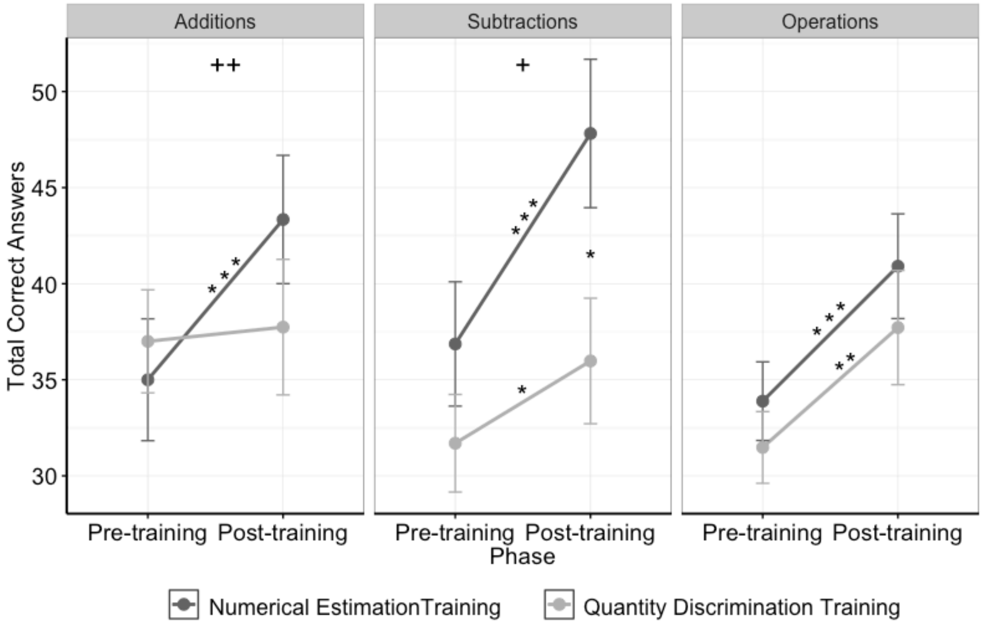


Figure 9. *Total correct answers in the Symbolic Tests in the Pre- and Post-training phases, for the Quantity Discrimination and Numerical Estimation Training conditions. Asterisks indicate significant post hoc tests. Plus signs indicate significant interactions, with number of symbols indicating the P-values of the tests (* p < .05, ** p < .01, *** p < .001, + p < .05, ++ p < .01).*

4. Discussion

Even the most basic calculations that we make in grade school (e.g., 2 + 2 =) involve a complex orchestration of mental abilities. An incomplete list of these abilities includes: different core systems of numbers, such as the Approximate Number System (Brannon, 2006; Feigenson et al., 2004; Halberda & Feigenson, 2008; Libertus & Brannon, 2009; Pica et al., 2004; Xu & Spelke, 2000) or our precise sense of very small quantities (D. C. Hyde & Spelke, 2011; Mou & VanMarle, 2014; Trick & Pylyshyn, 1993, 1994); special systems of notations for numbers, such as the Arabic digit code (Dehaene, 2001); words in a natural language expressing number, such as quantifiers for numerical expressions in English (Dehaene, 1992); and learned computational techniques, such as those performed with paper-and pencil (Ferres-Forga, Bonatti, et al., 2017; Nosworthy et al., 2013). Our increased awareness of the mental complexity of elementary mathematical operation offers exciting possibilities to devise novel ways to improve children's abilities to succeed in their first, crucial, years of mathematical education.

In Ferres-Forga *et al.* (2017), we explored the effects of a prolonged training of quantity discrimination in the ANS and its impact on children's school math performance. We found that such ANS training improves math performance in low-achieving children, and in particular, that it improves the comprehension of how a result varies as a function of the operations to be performed on the operands (addition, subtraction or multiplication). A possible explanation is that a prolonged training of discrimination with the ANS gives rise to a better understanding of how operations change quantities – although, in Ferres-Forga et al (2017) quantity discrimination training did not improve the children's performance with explicit addition and subtraction problems. In the present study, we found slightly larger benefits for quantity discrimination training, at least qualitatively. We again found that 3 weeks of Quantity Discrimination training led to improvements in understanding Operations – in this case for the group of children as a whole and not just for the low-achievers. We also found some improvement in Subtractions, but no improvements in Additions. It may be that Quantity Discrimination training will be most effective for younger and lower-achieving students. Further research is needed. What we can conclude, though, is that the potential benefits of Quantity Discrimination training are not as far-reaching as one might hope.

In the present study, we also investigated a second form of training – Numerical Estimation Training. This training focuses on a fundamental aspect of the basic understanding of numbers: the relation between their approximate quantities in the ANS and the exact cardinalities that refer to them, expressed as digits. In order to improve children's understanding of such relations, we devised a novel training regime: the Digits game. Its aim is to directly train the relation between a digit and the quantity it represents, providing the link between the ability to approximate the quantity of items without counting them and the symbolic language of mathematics. Notice that this link is not taught by the traditional Quantity Discrimination Training – which only focuses on comparing one approximate quantity with another, with no link to exact digits. Our intuition was that linking approximate quantities to digits would form a more robust type of training for transfer to school arithmetic.

We found that, over and above any potential improvement caused by ANS training, Numerical Estimation Training improves children's abilities to correctly solve exact additions and subtractions. For these computational abilities, training the link between Arabic digits and their nonsymbolic quantity representations proved to be beneficial above and beyond a highly comparable ANS Quantity Discrimination Training. Both training regimes had similar effects in the understanding of Operations. It is worth noticing that Operations performance requires no exact numerical accuracy, whereas the Additions and Subtractions Tests require formulating exact answers. Focusing children's attention on the exactness of the relation between digits and their referred quantities may better equip them to understand whether a calculation requiring an exact answer has indeed given the expected results. This better sense of exactness may build upon other positive effects of training provided by both training regimes.

The possible importance of building a mapping between quantity representations (e.g., the ANS) and our numerically discrete representations (e.g., digits, number words, number line) for math performance has been explored by several authors. Mappings from number words to quantities seem to emerge as children begin to understand counting (from 3-5 years of age) (Odic, Le Corre, et al., 2015), and mappings in the reverse direction, from quantities to number words, may emerge later (e.g., around 5-7 years of age) (Le Corre & Carey, 2007; Odic, Le Corre, et al., 2015). Both of these directions could help to support a sense of the relative size of numbers (e.g., how big is 17). And, this sense of “how big is 17” seems to also aid in children’s understanding of the linearly ordered number line emerging during the early elementary school years (Laski & Siegler, 2007; Siegler & Robinson, 1982). The ordering of digits may be an important factor in early elementary school, although it may not directly correlate with arithmetic achievement (Vogel, Remark, & Ansari, 2014). We showed that a training aimed at giving children a better grasp of the quantities associated to digits helps their mathematical abilities. Arguably, a better understanding of cardinalities may also improve children's grasp of ordinal relations among digits. And, later in life, an understanding of the ordering of digits may play a crucial role in linking our ANS and our math abilities, though the role of the ANS in such digit ordering tasks is hotly debated (Goffin & Ansari, 2016; Lyons & Beilock, 2013; Turconi, Jemel, Rossion, & Seron, 2004). At present, the conclusion from these literatures is that establishing a proper mapping between numerical symbols and their quantities may be important in order to improve mathematical abilities, but that much more research, across ages and across tasks is needed. In the current study, we sought to explicitly train the mapping from quantity to digits and we found significant improvements in Additions, Subtractions, and Operations performance on 7-8 year olds.

While the structure of the current study does not allow us to pinpoint the exact origin of the advantage that Numerical Estimation Training provided, it is interesting to speculate on why it may be effective. One possibility is that it may induce a better calibration of the meaning of digits. Numerical Estimation Training may cement anchor points of reference bridging the gap between the digit language and the world of quantities at some points of the numerical sequence (Izard & Dehaene, 2008). Recall that our Numerical Estimation Training provided increasing guidance to the correct number across the runs each day. And, because we maintained the spacing of answers across all target numbers (i.e., 1-21), answers at the highest level of difficulty (i.e., span 2) were very difficult and exposed children to a teaching signal relevant to each trial (e.g., *the answer is either 20, 21 or 22*), together with the passive trials in which the exact mapping between digits and quantities was also stressed. These factors may have been crucial for setting up correct anchors and training the children, providing them with a better understanding of the relation between digits and quantities. This deceptively simple form of calibration may have cascading effects. It may allow children to better exploit both symbolic and the nonsymbolic representations automatically created while dealing with numbers and quantities, essentially acting as a lever to multiply the efficiency of other pre-existing representations such as those generated by the ANS.

A second, non-exclusive possibility, is that Numerical Estimation Training may foster the process of the acquisition of numerical language. The relation between digits and quantities is essentially arbitrary (e.g., "three" & 3 $\rightarrow$ {x, y, z}). The acquisition of the language of numbers is a developmentally protracted process, not unlike the acquisition of another arbitrary but fundamental code such as reading. Mapping from digits or words to quantities can happen in either direction and at any moment, and children must understand the arbitrary nature of these mappings as well as their underlying regularities. Solidifying the understanding of the relation between digits and quantities could be as important as fostering phonological awareness in reading acquisition (Vanbinst, Ansari, Ghesquière, & Smedt, 2016). Neuropsychological data on the circuits involved in number processing may provide us some understanding of what such learning may do to the brain. Our approximate number system is associated with the intraparietal region of both hemispheres (Izard et al., 2008) while the match between quantities and digits recruits a distinct left-hemisphere circuit associated with linguistic symbols. The functional specialization of the left parietal circuit develops with schooling experience across a wide age span (Ansari & Dhital, 2006; Pinel &

Dehaene, 2010; Rivera et al., 2005). It is not impossible that procedures such as Numerical Estimation Training reinforce the establishment of a common and better functioning integrated network, precisely at the very onset of children's protracted period of exposure to mathematical education. Further research is needed to explore such hypotheses.

In conclusion, the current results, coupled with Ferres-Forga *et al.'s* (2017) training study, suggest that, while training based on Quantity Discrimination with the Approximate Number System could help children who are struggling in mathematics to better understand the basic arithmetic operations, in order to bring about an overall improvement in children's ability to successfully complete exact arithmetic calculations improving their understanding of the meanings of digits may be crucial. We submit that the educational systems could overestimate 7-8 year olds' comprehension of this basic aspect of mathematical language. An appropriate training of mapping from quantity to digit may complement standard school teaching routines, in an effortless and playful way, potentially generating long-lasting benefits in children's mathematical abilities and self-confidence, starting from these crucial founding years of their initiation to mathematical concepts.

Acknowledgments

The research was supported by a grant from the Catalan Government (SGR 2009-1521) to the Grup de Recerca en Neurociència Cognitiva (GRNC), MINECO PSI2016-75910-P to L.L.B., and a McDonnell Scholar Award to J.H.

Special thanks to Sònia Sas, general director of Hamelin Laie International School, for giving access to the participants and the school facilities essential to conduct this research. We thank the teachers Elvira Cabrespina and Mar Moreno for their collaboration with the experimenter during the completion of the research. We thank Miguel Burgaleta for helpful data analysis discussions.

References

Abreu-Mendoza, R. A., Soto-Alba, E. E., & Arias-Trejo, N. (2013). Area vs. density: influence of visual variables and cardinality knowledge in early number comparison. *Frontiers in Psychology*, *4*, 805. http://doi.org/10.3389/fpsyg.2013.00805

Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences*, 201603205. http://doi.org/10.1073/pnas.1603205113

Ansari, D., & Dhital, B. (2006). Age-related Changes in the Activation of the Intraparietal Sulcus during Nonsymbolic Magnitude Processing: An Event-related Functional Magnetic Resonance Imaging Study. *Journal of Cognitive Neuroscience*, *18*(11), 1820–1828. http://doi.org/10.1162/jocn.2006.18.11.1820

Bonny, J. W., & Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: evidence from the preschool years. *Journal of Experimental Child Psychology*, *114*(3), 375–88. http://doi.org/10.1016/j.jecp.2012.09.015

Brannon, E. M. (2006). The representation of numerical magnitude. *Current Opinion in Neurobiology*, *16*(2), 222–229. http://doi.org/10.1016/j.conb.2006.03.002

Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences*, *14*(12), 534–41. http://doi.org/10.1016/j.tics.2010.09.007

Carey, S. (2004). *Bootstrapping & the origin of concepts*. *Daedalus* (Vol. 133). http://doi.org/10.1162/001152604772746701

Cerda, G., Pérez, C., Navarro, J. I., Aguilar, M., Casas, J. a, & Aragón, E. (2015). Explanatory model of emotional-cognitive variables in school mathematics performance: a longitudinal study in primary school. *Frontiers in Psychology*, *6*(September), 1363. http://doi.org/10.3389/fpsyg.2015.01363

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, *148*, 163–172. http://doi.org/10.1016/j.actpsy.2014.01.016

Clayton, S., Gilmore, C., & Inglis, M. (2015). Dot comparison stimuli are not all alike: the effect of different visual controls on ANS measurement. *Acta Psychologica*, *161*, 177–84. http://doi.org/10.1016/j.actpsy.2015.09.007

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin {&} Review*, *8*(4), 698–707. http://doi.org/10.3758/BF03196206

Coubart, A., Izard, V., Spelke, E. S., Marie, J., & Streri, A. (2014). Dissociation between small and large numerosities in newborn infants. *Developmental Science*, *17*(1), 11–22. http://doi.org/10.1111/desc.12108

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1–2), 1–42. http://doi.org/10.1016/0010-0277(92)90049-N

Dehaene, S. (2001). Précis of The Number Sense. *Mind & Language*, *16*(1), 16–36. http://doi.org/10.1111/1468-0017.00154

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In P. Haggard & Y. Rossetti (Eds.), *Attention & Performance XXII. Sensori-motor foundations of higher cognition.* (pp. 527–574). Cambridge, Mass.: Harvard

Universitiy Press. http://doi.org/10.1093/acprof:oso/9780199231447.003.0024

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 314–326. http://doi.org/10.1037/0278-7393.21.2.314

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361. http://doi.org/10.1016/S0166-2236(98)01263-6

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 626–641. http://doi.org/10.1037/0096-1523.16.3.626

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*(1–2), 1–37. http://doi.org/10.1016/S0010-0277(00)00123-2

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., … Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), 597–600. http://doi.org/10.1038/26967

Demaray, M. K., & Elliot, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, *13*(1), 8–24. http://doi.org/10.1037/h0088969

den Heyer, K., & Briand, K. (1986). Priming single digit numbers: Automatic spreading activation dissipates as a function of semantic distance. *The American Journal of Psychology*, *99*(3), 315–340. http://doi.org/10.2307/1422488

Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. http://doi.org/10.1016/j.tics.2004.05.002

Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links Between the Intuitive Sense of Number and Formal Mathematics Ability. *Child Development Perspectives*, *7*(2), 74–79. http://doi.org/10.1111/cdep.12019

Ferres-Forga, N., Bonatti, L. L., & Halberda, J. (2017). One-Month of Approximate Number System Training Improves Symbolic Mathematics Performance, But Only In Low-Performing Children. *Submitted*.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequence is spatially organized. *Cognition*, *87*, B87–B95. http://doi.org/10.1016/S0010-9452(08)70938-9

Gevers, W., Reynvoet, B., & Fias, W. (2004). The Mental Representation of Ordinal Sequences is Spatially Organised: Evidence from Days of the Week. *Cortex*, *40*(1), 171–172. http://doi.org/10.1016/S0010-9452(08)70938-9

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology*, *76*(2), 104–122. http://doi.org/10.1006/jecp.2000.2564

Goffin, C., & Ansari, D. (2016). Beyond magnitude: Judging ordinality of symbolic number is unrelated to magnitude comparison and independently relates to individual differences in arithmetic. *Cognition*, *150*, 68–76. http://doi.org/10.1016/j.cognition.2016.01.018

Gordon, P. (2004). Numerical Cognition Without Words: Evidence from Amazonia. *Science*, *306*(5695), 496–499. http://doi.org/10.1126/science.1094492

Guillaume, M., Gevers, W., & Content, A. (2016). Assessing the Approximate Number System: no relation between numerical comparison and estimation tasks. *Psychological Research*, *80*(2), 248–58. http://doi.org/10.1007/s00426-015-0657-x

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), 1457–65. http://doi.org/10.1037/a0012682

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(28), 11116–20. http://doi.org/10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–8. http://doi.org/10.1038/nature07246

He, L., Zhou, K., Zhou, T., He, S., & Chen, L. (2015). Topology-defined units in numerosity perception. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41), E5647-55. http://doi.org/10.1073/pnas.1512408112

He, Y., Zhou, X., Shi, D., Song, H., Zhang, H., & Shi, J. (2016). New Evidence on Causal Relationship between Approximate Number System (ANS) Acuity and Arithmetic Ability in Elementary-School Students: A Longitudinal Cross-Lagged Analysis. *Frontiers in Psychology*, *7*(July), 1–8. http://doi.org/10.3389/fpsyg.2016.01052

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, *10*(4), 389–395. http://doi.org/10.3758/BF03202431

Hinrichs, J. V, Yurko, D. S., & Hu, J. (1981). Two-digit number comparison: Use of place information. *Journal of Experimental*

Psychology: Human Perception and Performance, *7*(4), 890–901. http://doi.org/10.1037/0096-1523.7.4.890

Hollingsworth, W. H., Simmons, J. P., Coates, T. R., & Cross, H. A. (1991). Perceived numerosity as a function of array number, speed of array development, and density of array Items. *Bulletin of the Psychonomic Society*, *29*(5), 448–450. http://doi.org/10.3758/BF03333967

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, *131*(1), 92–107. http://doi.org/10.1016/j.cognition.2013.12.007

Hyde, D. C., & Spelke, E. S. (2011). Neural signatures of number processing in human infants: evidence for two core systems underlying numerical cognition. *Developmental Science*, *14*(2), 360–71. http://doi.org/10.1111/j.1467-7687.2010.00987.x

Izard, V., Dehaene-Lambertz, G., & Dehaene, S. (2008). Distinct cerebral pathways for object identity and number in human infants. *PLoS Biology*, *6*(2), 0275–0285. http://doi.org/10.1371/journal.pbio.0060011

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–47. http://doi.org/10.1016/j.cognition.2007.06.004

Kemp, S. (1984). Estimating the Sizes of Sports Crowds. *Perceptual and Motor Skills*, *59*(3), 723–729. http://doi.org/10.2466/pms.1984.59.3.723

Krueger, L. E. (1982). Single judgments of numerosity. *Perception {&} Psychophysics*, *31*(2), 175–182. http://doi.org/10.3758/BF03206218

Krueger, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. *Behavioral and Brain Sciences*, *12*(2), 251–267. http://doi.org/10.1017/S0140525X0004855X

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development*, *78*(6), 1723–1743. http://doi.org/10.1111/j.1467-8624.2007.01087.x

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395–438. http://doi.org/10.1016/j.cognition.2006.10.005

Leibovich, T., & Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Frontiers in Psychology*, *4*, 375. http://doi.org/10.3389/fpsyg.2013.00375

Lemer, C., Dehaene, S., Spelke, E. S., & Cohen, L. (2003). Approximate quantities and exact number words: dissociable systems. *Neuropsychologia*, *41*(14), 1942–58. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/14572527

Libertus, M. E., & Brannon, E. M. (2009). Behavioral and Neural Basis of Number Sense in Infancy. *Current Directions in Psychological Science*, *18*(6), 346–351. http://doi.org/10.1111/j.1467-8721.2009.01665.x

Libertus, M. E., & Brannon, E. M. (2010). Stable individual differences in number discrimination in infancy. *Developmental Science*, *13*(6), 900–6. http://doi.org/10.1111/j.1467-7687.2009.00948.x

Libertus, M. E., Feigenson, L., & Halberda, J. (2013a). Is Approximate Number Precision a Stable Predictor of Math Ability? *Learning and Individual Differences*, *25*, 126–133. http://doi.org/10.1016/j.lindif.2013.02.001

Libertus, M. E., Feigenson, L., & Halberda, J. (2013b). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *Journal of Experimental Child Psychology*, *116*(4), 829–38. http://doi.org/10.1016/j.jecp.2013.08.003

Lyons, I. M., & Beilock, S. L. (2013). Ordinality and the nature of symbolic numbers. *The Journal of Neuroscience*, *33*(43), 17052–61. http://doi.org/10.1523/JNEUROSCI.1775-13.2013

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011a). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development*, *82*(4), 1224–1237. http://doi.org/10.1111/j.1467-8624.2011.01608.x

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011b). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE, 6*(9), 1–9. http://doi.org/10.1371/journal.pone.0023749

McCrink, K., Spelke, E. S., Dehaene, S., & Pica, P. (2013). Non-symbolic halving in an Amazonian indigene group. *Developmental Science*, *16*(3), 451–62. http://doi.org/10.1111/desc.12037

Moll, K., Snowling, M. J., Göbel, S. M., & Hulme, C. (2015). Early language and executive skills predict variations in number and arithmetic skills in children at family-risk of dyslexia and typically developing controls. *Learning and Instruction*, *38*, 53–62. http://doi.org/10.1016/j.learninstruc.2015.03.004

Mou, Y., & VanMarle, K. (2014). Two core systems of numerical representation in infants. *Developmental Review*, *34*(1), 1–25. http://doi.org/10.1016/j.dr.2013.11.001

Moyer, R. S., & Landauer, T. K. (1967). Time required for Judgements of Numerical Inequality. *Nature, 215*(5109), 1519–1520. http://doi.org/10.1038/2151519a0

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology, 103*(4), 490–502. http://doi.org/10.1016/j.jecp.2009.02.003

Naccache, L., & Dehaene, S. (2001a). The Priming Method: Imaging Unconscious Repetition Priming Reveals an Abstract Representation of Number in the Parietal Lobes. *Cerebral Cortex*, *11*(10), 966–974. http://doi.org/10.1093/cercor/11.10.966

Naccache, L., & Dehaene, S. (2001b). Unconscious semantic priming extends to novel unseen stimuli. *Cognition*, *80*(3), 223–237. http://doi.org/10.1016/S0010-0277(00)00139-6

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PloS One*, *8*(7), e67918. http://doi.org/10.1371/journal.pone.0067918

Núñez-Peña, M. I., & Suárez-Pellicioni, M. (2014). Less precise representation of numerical magnitude in high math-anxious individuals: An ERP study of the size and distance effects. *Biological Psychology*, *103*, 176–183. http://doi.org/10.1016/j.biopsycho.2014.09.004

Odic, D., Hock, H., & Halberda, J. (2012). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology. General*, *142*(1), 255–65. http://doi.org/10.1037/a0030825

Odic, D., Im, H. Y., Eisinger, R., Ly, R., & Halberda, J. (2015). PsiMLE: A maximum-likelihood estimation approach to estimating psychophysical scaling and variability more reliably, efficiently, and flexibly. *Behavior Research Methods*. http://doi.org/10.3758/s13428-015-0600-5

Odic, D., Le Corre, M., & Halberda, J. (2015). Children's mappings between number words and the approximate number system. *Cognition*, *138*, 102–21. http://doi.org/10.1016/j.cognition.2015.01.008

Park, J., & Starns, J. J. (2015). The approximate number system acuity redefined: A diffusion model approach. *Frontiers in Psychology*, *6*(DEC), 1–10. http://doi.org/10.3389/fpsyg.2015.01955

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555. http://doi.org/10.1016/j.neuron.2004.10.014

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science (New York, N.Y.)*, *306*(5695), 499–503. http://doi.org/10.1126/science.1102085

Pinel, P., & Dehaene, S. (2010). Beyond Hemispheric Dominance: Brain Regions Underlying the Joint Lateralization of Language and Arithmetic to the Left Hemisphere. *Journal of Cognitive Neuroscience*, *22*(1), 48–66. http://doi.org/10.1162/jocn.2009.21184

Pinel, P., Dehaene, S., Rivière, D., & LeBihan, D. (2001). Modulation of Parietal Activation by Semantic Distance in a Number Comparison Task. *NeuroImage*, *14*(5), 1013–1026. http://doi.org/10.1006/nimg.2001.0913

Pletzer, B., Kronbichler, M., Nuerk, H. C., & Kerschbaum, H. H. (2015). Mathematics anxiety reduces default mode network deactivation in response to numerical tasks. *Frontiers in Human Neuroscience*, *9*(April), 1–12. http://doi.org/10.3389/fnhum.2015.00202

Purpura, D. J., & Logan, J. A. R. (2015). The nonlinear relations of the approximate number system and mathematical language to early mathematics development. *Developmental Psychology*, *51*(12), 1717–24. http://doi.org/10.1037/dev0000055

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does Subitizing Reflect Numerical Estimation?

Psychological Science, 19(6), 607–614. http://doi.org/10.1111/j.1467-9280.2008.02130.x

Reynvoet, B., & Ratinckx, E. (2004). Hemispheric differences between left and right number representations: Effects of conscious and unconscious priming. *Neuropsychologia*, *42*(6), 713–726. http://doi.org/10.1016/j.neuropsychologia.2003.11.013

Rivera, S. M., Reiss, A. L., Eckert, M. A., & Menon, V. (2005). Developmental changes in mental arithmetic: Evidence for increased functional specialization in the left inferior parietal cortex. *Cerebral Cortex*, *15*(11), 1779–1790. http://doi.org/10.1093/cercor/bhi055

Rouder, J. N., & Geary, D. C. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science*, *17*(4), 525–36. http://doi.org/10.1111/desc.12166

Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*(3), 662–674. http://doi.org/10.1016/j.cognition.2008.05.007

Shusterman, A., Slusser, E., Halberda, J., & Odic, D. (2016). Acquisition of the cardinal principle coincides with improvement in approximate number system acuity in preschoolers. *PLoS One*, *11*(4), e0153072. http://doi.org/10.1371/journal.pone.0153072

Siegler, R. S., & Robinson, M. (1982). The Development of Numerical Understandings. *Advances in Child Development and Behavior*, *16*(C), 241–312. http://doi.org/10.1016/S0065-2407(08)60072-5

Simzar, R. M., Domina, T., & Tran, C. (2016). Eighth Grade Algebra Course Placement and Student Motivation for Mathematics. *HHS Public Access*, *2*(1), 1–26. http://doi.org/10.1177/2332858415625227

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18116–20. http://doi.org/10.1073/pnas.1302751110

Stoet, G., Bailey, D. H., Moore, A. M., & Geary, D. C. (2016). Countries with higher levels of gender equality show larger national sex differences in mathematics anxiety and relatively lower parental mathematics valuation for girls. *Plos One*, *11*, e0153857. http://doi.org/10.1371/journal.pone.0153857

Sullivan, J., & Barner, D. (2014). The development of structural analogy in number-line estimation. *Journal of Experimental Child Psychology*, *128*, 171–89. http://doi.org/10.1016/j.jecp.2014.07.004

Thien, L. M., & Ong, M. Y. (2015). Malaysian and Singaporean students' affective characteristics and mathematics performance: evidence from PISA 2012. *SpringerPlus*, *4*(1), 563. http://doi.org/10.1186/s40064-015-1358-z

Tournaki, N. (2003). Effect of Student Characteristics on Teachers' Predictions of Student Success. *The Journal of Educational Research*, *96*(5), 310–319. http://doi.org/10.1080/00220670309597643

Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(2), 331–351. http://doi.org/10.1037/0096-1523.19.2.331

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*(1), 80–102. http://doi.org/10.1037/0033-295X.101.1.80

Turconi, E., Jemel, B., Rossion, B., & Seron, X. (2004). Electrophysiological evidence for differential processing of numerical quantity and order in humans. *Cognitive Brain Research*, *21*(1), 22–38. http://doi.org/10.1016/j.cogbrainres.2004.05.003

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 166–179. http://doi.org/10.1037/0278-7393.18.1.166

van Oeffelen, M. P., & Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics*, *32*(2), 163–170. http://doi.org/10.3758/BF03204275

Vanbinst, K., Ansari, D., Ghesquière, P., & Smedt, B. De. (2016). Symbolic numerical magnitude proceing is as important to arithmetic as phonological awarene is to reading. *PLoS ONE*, *11*(3), 1–11. http://doi.org/10.1371/journal.pone.0151045

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., & Chang, A. (2014). Deconstructing Building Blocks: Preschoolers' Spatial Assembly Performance Relates to Early Mathematical Skills. *Child Development*, *85*(3), 1062–1076. http://doi.org/10.1111/cdev.12165

Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition*, *119*(1), 10–22. http://doi.org/10.1016/j.cognition.2010.11.014

Wang, J., Odic, D., Halberda, J., & Feigenson, L. (2016). Changing the precision of preschoolers' approximate number system representations changes their symbolic math performance. *Journal of Experimental Child Psychology*, *147*, 82–99. http://doi.org/10.1016/j.jecp.2016.03.002

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193. http://doi.org/10.1016/0010-0277(90)90003-3

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*(2), 220–251. http://doi.org/10.1016/0010-0285(92)90008-P

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. http://doi.org/10.1016/S0010-0277(99)00066-9

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science*, *8*(1), 88–101. http://doi.org/10.1111/j.1467-7687.2005.00395.x

4. EXPERIMENTAL SECTION 3

4.1 Introduction

Our intuitive sense of number, generally described as the Approximate Number System (ANS), is the basis for a non-linguistic ability to approximate quantity representations without the need for symbolic numbers or counting skills (Butterworth, 1999; Dehaene, 1997; Halberda, Mazzocco, & Feigenson, 2008). We share such system with other animals (Dehaene, Dehaene-Lambertz, et al., 1998; Elena et al., 2016) and preverbal infants (Brannon, 2006; Feigenson et al., 2004; Hyde, 2011; Hyde & Spelke, 2011; Libertus & Brannon, 2009; Xu & Spelke, 2000; Xu et al., 2005; Izard et al., 2008). Even newborn infants seem to have numerical representations (Antell & Keating, 1983; Coubart et al., 2014; de Hevia et al., 2014; Izard et al., 2009).

The way in which ANS represents quantity information is such that smaller quantities are represented more precisely than larger quantities following Weber's Law (Dehaene, 1997; Feigenson, Dehaene & Spelke, 2004). The Weber fraction can thus be understood as an index of ANS acuity (Dehaene, 1997). The precision of the ANS increases with cognitive development, indicating some flexibility of this capacity (Halberda & Feigenson, 2008). Some researches indicate that ANS precision is trainable (Knoll et al., 2016; Obersteiner et al., 2013; Park et al., 2016; Park & Brannon, 2014).

Several lines of research suggest that ANS acuity relates with symbolic math performance (Feigenson et al., 2013) throughout development and into adulthood (Halberda, Mazzocco & Feigenson, 2008; Halberda et al., 2012)(Starr, Libertus & Brannon, 2014; Jordan, Kaplan, Olah & Locuniak, 2006).

In Studies 1 and 2 (Chapters 2 and 3), we trained the ANS with a quantity discrimination task by playing the Panamath computer program during a period of three weeks. Participants were 7-8-year-old children. We verified that our ANS training showed a ratio-dependent performance predicted by Weber's law. We could also verify that ANS precision was trainable in a short period and in a longer period since the efficiency in the performance of the quantity discrimination task increased significantly within each day and over the course of the training. We could also see the relation between ANS acuity and symbolic math performance (Figure 5, Study 1, Chapter 2). Finally, the training transferred improvements to the symbolic math test (Figures 1a and 1b, Chapters 2 and 3). However, in both studies, benefits were mainly transferred to the operations test in which children had to find the operation that accomplishes the equality (3 ☐ 5 = 15; addition, subtraction or multiplication), but any numerical answer is required (Figures 11-12, Study 1, Chapter 2; Figure 9, Study 2, Chapter 3).

The Approximate Number System is not sufficient to compute exact operations with numbers and to progress to more advanced mathematics; an accurate mathematical language is needed (Bonny & Lourenco, 2013; Butterworth, 2010; Dehaene, 2001; Gordon, 2004; Lemer et al., 2003; McCrink et al., 2013; M. Piazza, Pica, Izard, Spelke, & Dehaene, 2013; Pica et al., 2004). Arabic digits are a precise representation of quantities and understanding its cardinality is crucial for the improvement of arithmetic skills. For example, Mundy & Gilmore (2009) found that in 6-8 year-old children the accuracy in mapping Arabic digits and nonsymbolic representations was related to mathematics achievement.

In Study 2 (Chapter 3, Ferres-Forga, Halberda, & Bonatti, 2017) we independently trained the ANS with the same quantity discrimination task as in Study 1, and the mapping between digits and quantities with a Numerical Estimation task by asking children to play with the Digits program, in a three-week computer-trained regime. We verified that the precision of the mapping between digits and quantities could be trained (Figure 8, Study 2, Chapter 3). We found that strengthening the digits-quantity mapping results in a generalized improvement of 7-year-olds' math competence. Such improvement adds to those that a regime based on training ANS can provide. Thus, we propose that training the DQR helps children to become more proficient in exact mathematics.

The ANS and the accurate symbolic math abilities rely on partially separate cerebral circuits with a differential lateralization: left inferior parietal cortex appears to be specialized for symbolic numbers processing, while right superior parietal lobule for nonsymbolic sets of items (ANS) (Dehaene & Cohen, 1991, 1997; Sokolowski et al., 2017). In addition, symbolic math abilities are more dependent on learning than ANS abilities. We would like to know how they are related and how this relation changes across ages.

With this background, we wanted to extend our research in several aspects:

1) We wanted to confirm that the accuracy in ANS increases with age.

2) We wanted to know how stable the relation between ANS accuracy and school math performance evolves across years. This question is important because we did find a relation between ANS and symbolic math in 7-8 year olds, but this relation was not found when an exact numerical answer was required.

3) We wanted to verify if the accuracy in the mapping between Arabic digits and quantities improves with age.

4) We wanted to know if the relation between the digit-quantity mapping and the symbolic math performance that we found in 7-8 years old could be also found in older students.

5) We wanted to have within-participant measures of ANS accuracy and digit-quantity mapping accuracy, so that we could accurately study the correlations of this two abilities, and across ages. Indeed, in Studies 1 and 2, participants belonged to one or the other training, and it was not possible to do otherwise given our training design. For this reason, we do not have data referring to the same

students about the two kinds of accuracies. Information about this point appear to us to be crucial.

These were the objectives of our Study 3. Our math tests have proven to be an appropriate tool for assessing the level of arithmetic of second grade children (7-8 year olds) before and after training. However, since we wanted to extend the research to older students, we needed a standard and age-related tool to measure students' math competence. To this purpose, we decided to use the school math marks of all students who participated in the study. The School kindly provided us with these data.

Our previous researches were centered on second-grade students, that is on 7-to-8 year olds. In the present research, we could examine the records of 529 students aged between 8 and 16 years, going from third to tenth grade. Participants performed a one-hour session in which they played two games: a Numerical Estimation task (Digits program) and a Quantity Discrimination task (Panamath program). The order of the games was counterbalanced between participants. In addition, we could consult their school math marks at their second and third assessment period, corresponding to the second and third quarter of the school year (2^{nd} Evaluation and 3^{rd} Evaluation hereafter). The test session was administered at the beginning of the third quarter. Thus, the session was at the beginning of the third quarter. In this way, the school marks at 2^{nd} and 3^{rd} Evaluations corresponded to two equivalent periods before and after the session. In addition, using the two math scores closest to the session, instead of one, gave us more stability as we avoided some transient and unrepresentative fluctuations of some students. Therefore, we have used the school math marks of 2^{nd} and 3^{rd} Evaluations to strengthen our conclusions about the correlations we have found.

We report our detailed analyses below. In summary, our results show that ANS accuracy was affected by the ratio of the set sizes to be compared and was predicted by Weber's law in all ages. This accuracy increased with age, from 8 to 16 years. Furthermore, the accuracy in the mapping between Arabic digits and quantities showed size, distance and position effects. It also increased with age. However, the relation between accuracies in ANS and the mapping between digits and quantities was not straightforward, as it turned out to be age-dependent. A correlation existed at grades 4th to 8th, but there was not such correlation among our youngest participants (3rd grade), nor among our oldest students (9th to 10th). Taking each task separately, in both tasks we found a positive correlation between percentage correct and response time. That is, the longer it took participants to answer, the greater their percentage of correct answers was. Turning to the relation between our tasks and school math scores, we found that the percentage of correct answers in the numerical estimation task (mapping digits-quantity) positively correlates with math school grades, both for 2nd and 3rd Evaluations, for a continuous and wide range of students from grade 3rd to grade 7th (8-to-13 years). For the quantity discrimination task (ANS), the correlation with school math in both evaluations was only found for 6th, 7th, and 10th grades, making the interpretation of this relation more difficult.

4.2 Methods

4.2.1 Participants

Students of the Hamelin International Laie School (http://www.hamelininternationallaie.com/school) participated in this study (n= 529; 257 girls; average age = 11 y 9 m, range 8 y 2 m to 16 y 9 m). They were pupils from third to tenth grade. Twenty-four classes were involved: four of 3rd grade (n = 92; 45 girls; average age = 8 y 9 m, range 8 y 2 m to 9 y 6 m); four of 4th grade (n = 86; 36 girls; average age = 9 y 9 m, range 9 y 1 m to 10 y 5 m); three of 5th grade (n = 62; 34 girls; average age = 10 y 10 m, range 9 y 6 m to 11 y 10 m); three of 6th grade (n = 79; 41 girls; average age = 11 y 11 m, range 11 y 1 m to 15 y 3 m), three of 7th grade (n = 72; 30 girls; average age = 12 y 10 m, range 12 y to 14 y 3 m), three of 8th grade (n = 58; 23 girls; average age = 13 y 9 m, range 13 y to 14 y 3 m), two of 9th grade (n = 41; 22 girls; average age = 14 y 10 m, range 14 y 3 m to 15 y 8 m); and two of 10th grade (n = 39; 26 girls; average age = 15 y 10 m, range 14 y 6 m to 16 y 9 m). The students mostly came from middle to high socioeconomic status families.

The study consisted of one one-hour session, conducted at the premises of the school. The session was integrated in the class schedule, during the computer technology class-time. Participants played two games: a Numerical Estimation task and a Quantity Discrimination task. The order was counterbalanced between participants. Four students were excluded because of impaired hand motor skills (n=1, 5th grade), impaired vision (n=1, 9th grade), comprehension problems (n=1, 6th grade) and behavioral problems (n=1, 7th grade).

Numerical Estimation task. We retained a total of 524 participants for analysis (254 girls). Of them, 516 completed 100% of the trials. Other students who did not completed the full set of trials (90% of the trials, n=1; 78% of the trials, n=7; 50% of the trials, n=1) were included in the analysis. One 4th grade student completed less than 8% of the trials and was excluded from further analyses. Of the 524 participants retained for analysis, 92 students (45 girls) were in 3rd grade, 85 students (35 girls) in 4th grade, 61students (33 girls) in 5th grade, 78 students (40 girls) in 6th grade, 71 students (30 girls) in 7th grade, 58 students (23 girls) in 8th grade, 40 students (22 girls) in 9th grade and 39 students (26 girls) in 10th grade.

Quantity Discrimination task. We retained a total of 503 participants for analysis (244 girls). Of them, 500 completed the 100% of the trials and 3 completed the 50% of the trials. Technical problems precluded some students from performing the test (n=22). Of the participants retained for analysis, 71 students (35 girls) were in 3rd grade, 86 students (36 girls) in 4th grade, 61 students (33 girls) in 5th grade, 77 students (39 girls) in 6th grade, 71 students (30 girls) in 7th grade, 58 students (23 girls) in 8th grade, 40 students (22 girls) in 9th grade and finally, 39 students (26 girls) belonged to 10th grade.

4.2.2 Materials

Computer Activities. Twenty-eight computers (model: clon PCs, Intel® Core™ i3-4170 CPU @ 3.70 GHz, 4GB RAM, 64-bit; monitor: 17” LCD 16/9 ASUS; operative system: Windows 7 Professional) were used for the session. The students wore headphones during training. While both activities (Quantity Discrimination and Numerical Estimation tasks) were new to them, students were already acquainted with the material and the computer classroom.

a) Quantity Discrimination Task

For this activity, the computer ran a modified version of the computer game Panamath (Halberda et al., 2008), written in Java SE6. In this version, the program would generate trials displaying collections of items contained inside two rectangles appearing on the sides of the screen. For example, twelve yellow dots could appear inside the left rectangle and six blue dots inside the right rectangle. The number of items within each rectangle was always between 5 and 21. The items were presented in seven different ratios (larger set/smaller set). The ratios could be 3, 2, 1.5, 1.25, 1.17, 1.14, and 1.1. For example, on a 3-ratio trial children could see 21 blue dots in the right side of the screen and 7 yellow dots in the left side. Smaller ratios correspond to more difficult trials. On each trial, the items were displayed for 1382 msec. A run was composed of 35 consecutive trials. Different collections of items were used on each run, so that the game maintained children's interest. For example, one run could present blue dots vs. yellow dots; another could present cars vs. bears; yet another could display birds vs. dogs.

To vary the relation between surface area and number during the session, Panamath game implemented two different models controlling object size: half of the trials were size-confounded (Figure 4.1a), and the other half were size-controlled (Figure 4.1b). In size-confounded trials, the average size of the items was equal for both sets, so that the cumulative surface area occupied by the objects was congruent with the number of objects. In size-controlled trials, the average size of the objects was smaller for the larger set, so that the ratio of the cumulative area occupied by the objects in each set was equated.

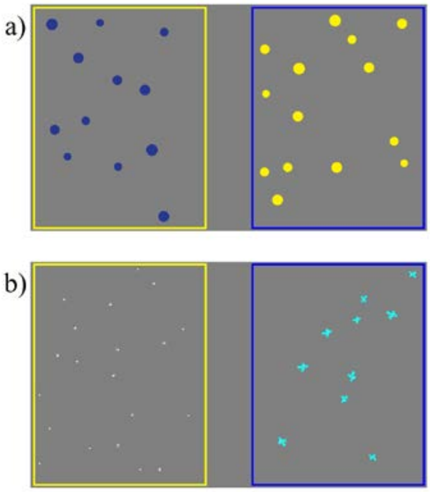


Figure 4.1. *Examples of the two models of objects size to control for surface area: size-confounded (a), size-controlled (b) used in the trials of this study.*

b) Numerical Estimation task

For this activity, the computer ran the "Digits" game, written in PsychoPy v1.83.01. The Digits game program generated two types of trials, the *passive trials* and the *active trials*. A run was composed by thirty-five consecutive trials; seven of them were passive trials and twenty-eight were active trials. In the passive trials, a collection of items was presented on screen while a prerecorded voice named the exact number of items. These trials lasted 1200 ms (Figure 4.2a). In the active trials, the program first presented a collection of items for 1 s, in silence (Figure 4.2a).

Then, the collection disappeared and three digits were presented on screen until the child clicked on one of them, ideally the one that represented the quantity of the previous collection of items (Figure 4.2b). Different collections of items were used in each run to keep students' interest (dots, tucks, dogs…).

The purpose of the passive trials was to provide participants opportunities to directly calibrate their estimation system before the active trials began (Izard & Dehaene, 2008; Krueger, 1989). Considering that these trials were passive, the program presented them in decreasing order of difficulty (from bigger to smaller sets of items), so that the easier passive trials were always presented last, thus increasing students' confidence before the active trials started.

Active trials were presented in increasing order of difficulty, a procedure which is known to facilitate learning (Odic et al., 2012; J. Wang et al., 2016). Because the distance between digits and the numerosity of the set presented in each trial are subject to known effects (respectively, the distance and size effects (Cordes et al., 2001; Dehaene, 2007; Dehaene, Dehaene-Lambertz, et al., 1998; van Oeffelen & Vos, 1982)) we manipulated these factors in the construction of the game. Thus, the distance between the correct choice (that is, the digit representing the exact number of items) and the distractor choices manipulated the distance effect. Distances ranged between -6 and 6 from the true value; the greater the distance between the digits, the easier the task of the participant. For example, if the target number was 15, a decision between 18, 21, and 15 (or +3, +6, and 0 distance from the correct choice) was easier than the decision between 15, 13, and 17 (0, +2, -2). Thus, throughout each set, the distances between the correct digit and the other two digits decreased in each run. For the first run, distances could be (0, 3, 6), (6, -3, 0), or (-3, 0, 3). We call this set *Span 6*. For the second run, distances could be (0, 2, 4), (-4, -2, 0), or (-2, 0, 2), or *Span 4*. Finally, for the third run distances could be (0, 1, 2), (-2, -1, 0), or (-1, 0, 1), or *Span 2*.

The size effect was manipulated by incrementing the size of the collections presented in active trials. Therefore, by manipulating distance and size, the difficulty of the game increased within each run. All set sizes, from 1 item to 21 items, had to be estimated in each of the three spans. Furthermore, the distances for possible answers were maintained irrespective of the target answer, rather than scaling the distractor answers relative to the true answer by some ratio (e.g., larger distances for the competing answers as the target value becomes larger). This was a conscious explicit choice. While scaling the answers is also possible, and might even be preferred in methods used to assess Approximate Number System precision rather than assessing the numerical estimation precision, here we wanted trial difficulty to increase with target number. This has the goal of providing greater guidance for larger quantities; for example, on a trial with target value 19, the three answer options on a difficult *Span 2* trial would all provide a strong guidance signal (e.g., 18, 19, 20) for helping to show participants the correct numerical region for an answer (e.g., helping to overcome an underestimation bias).

The position of the correct digit (left, middle, right) as well as its numerical relation with respect to the other two digits (the smallest, in the middle, the largest) were balanced across trials.

In both types of trials (passive and active), the collections of items ranged from 1 to 21, all with the same size and orientation.

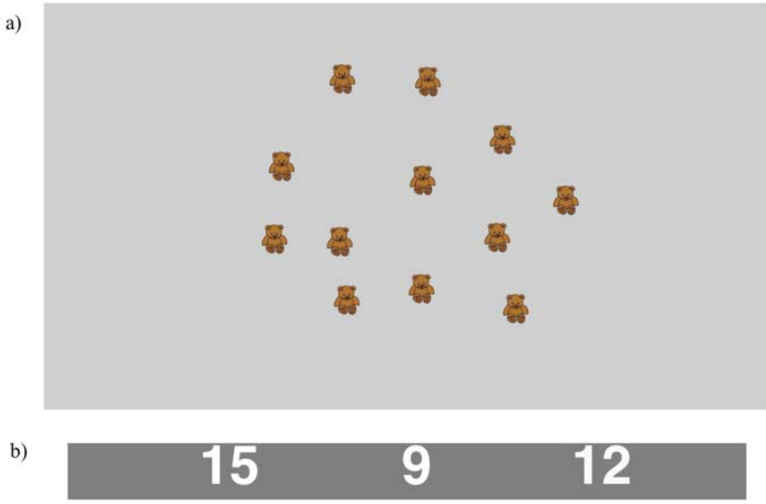


Figure 4.2. *Digits computer game. Example of a trial (a) that would be the same for either passive (i.e., with a spoken numerical label "twelve") or active trial (i.e., no verbal label), and (b) the response screen for the active trial. The collection of items (a) for an active trial would appear for 1 s, in silence, then, the collection would disappear and three answer options (b) would be presented on screen until the student clicked on one of them.*

4.2.3 Procedure

The activity was performed in one one-hour session at each of the 24 classes. All sessions were held in presence of the math teacher of the class and of the experimenter. After the experimenter gave them instructions, participants worked individually.

Students played two whole rounds for each game, Panamath (quantity discrimination task) and Digits (numerical estimation task). Each round consisted of three runs, each of which was formed by 35 trials. Each run took approximately 5 minutes. The difficulty of the game was reset at each round. After the session, students appeared to like both games.

a) Quantity Discrimination Task

To evaluate the approximate number system accuracy of the students they performed a Quantity Discrimination task by playing Panamath game. This game does not require any understanding of the relation between digits and quantities. On each trial, students saw a picture of two collections of items appear on either side of the screen. They needed to rapidly estimate which side had more items (Figure 4.1 a-b), typing their answers on a keyboard ("f" and "j" keys for left or right side respectively). Students could not count the items, because the onscreen presentation of the objects was too brief (≈1.3 seconds). Rather, they had to rapidly estimate which of the two sets in each side of the screen had more items. Feedback was provided after every response: a high-pitched beep indicated a correct answer while a low-pitched beep indicated an incorrect answer. Thirty-five consecutive trials formed a run. Always, the first five trials of each run presented the easiest ratio. Then, every five trials the game increased in difficulty, with the ratios becoming closer to 1 (without ever reaching 1), until the seven different ratios were presented. This procedure was implemented with the aim to increase participants' ANS precision and confidence over the course of the session, as observed in previous studies using brief interventions (Odic et al., 2012; J. Wang et al., 2016).

When introducing this Quantity Discrimination task, students were told that they would play a game where they would see some objects -- for example, blue and yellow dots --and would have to choose if there were more blue dots than yellow, or vice-versa. They were informed that two different sounds would provide them with feedback about the correctness of their answer. Students were also told that the game would increase in difficulty throughout each round and that they would play the game twice (two whole rounds). They were informed that both speed and accuracy were important.

b) Numerical Estimation Task

To assess students' knowledge required to map from quantities to the exactitude of the digits, participants performed a

Numerical Estimation task by playing Digits game. In this game, the thirty-five trials presented in each run differentiated into two types: passive and active trials. Each run started with seven passive trials (Figure 4.2a). Students need not to take any action during these trials. The following twenty-eight trials composing each run were active trials (Figures 4.2a and 4.2b, never presented simultaneously on the screen). The collection of items appearing on screen remained visible for a too short time for participants to be able to count the number of items. Rather, they could respond by giving their best guess. The three digits among which they had to choose remained on screen until the choice was made, by clicking the mouse on one of them. Feedback was provided after every response, with a high-pitch beep for correct answers and a low-pitch beep for incorrect answers, as in the quantity discrimination task. And, as explained above, the trial difficulty increased with every run in each round.

When introducing the Digits game, students were told they would play a game where they would first see a collection of objects for a very short time, while an audio recording would tell them how many objects were in the collection. They were told that they would have to pay attention to these trials, but not take any action. They were also informed that they would then see many trials where a collection would be shown for a short period, after which they would have to decide their numerosity by choosing one of three digits that would appear on screen immediately after the disappearance of the items. They were informed that two different sounds would provide them with feedback about the correctness of their answer. Students were also told that the difficulty of the game would increase throughout each round and that they would play twice the game (two whole rounds). They were informed that both speed and accuracy were important.

4.3 Results

Because of the complexity of the results, we summarize the main structure of this section. We first analyze the responses to the two tasks separately, in order to draw a profile of the tasks (Section 4.3.1 and 4.3.2) We will then turn our analysis to the relation between the tasks and participant's school math performance. Finally, we will study the relations among tasks. Appendix xx and yy will present the details of some of the analysis presented here.

4.3.1 Quantity Discrimination Task

We recall that in this task, participants (n= 503) briefly viewed pictures of two collections of items appearing on either side of the screen. For each picture, they had to rapidly estimate which side had more items. Students completed a total of six runs (35 trials each run) with the same structure. In them, the ratio between the two sets of items decreased every 5 trials, making the game to become more difficult within each run.

4.3.1.1 Percentage of correct answers. Size and distance effect

We checked that performance in the Quantity Discrimination task showed the main signature of the Approximate Number System (Feigenson et al., 2004; Libertus & Brannon, 2009; Starr, Libertus, & Brannon, 2013). To anticipate the main result of this part of our analysis, participants' accuracy at determining the bigger of two approximate numerosities decreased as the ratio between the numbers decreased. This ratio-dependence is predicted by Weber's law with a specific curve of percentage of correct answers as a function of ratio (Figure 4.3). Below, I report more detailed analyses.

Figure 4.3 presents the data from the Quantity Discrimination task from all grades together (8-16 year olds; n= 503), separated by the type of size control for the stimuli. The ratio-dependent performance curve is observed for both trial types. That is, as the numerical ratio between the two collections became easier (e.g., ratio 3 versus ratio 1.2) students' percentage of correct responses improved, regardless of the type of size control for the trial. The curves in Figure 4.3 are generated by fitting a model of Weber's law to the mean performance of student in each ratio for each size control type. That is, each student contributes equally to the curves, the curves are fit to the group means, and the error bars are ±SE for the group performance.

Participants chose the numerically greater collection well above chance as shown by planned *t*-tests for Size Confounded trials: t (502) = 83.2, p < 0.001; and for Size Controlled trials, t (502) = 80.4, p < 0.001). Students' performance exhibits the smooth curve of the Approximate Number System. Therefore, we suggest that even if size contributes somewhat to children's decisions, children's numerical decisions were likely based on the ANS.

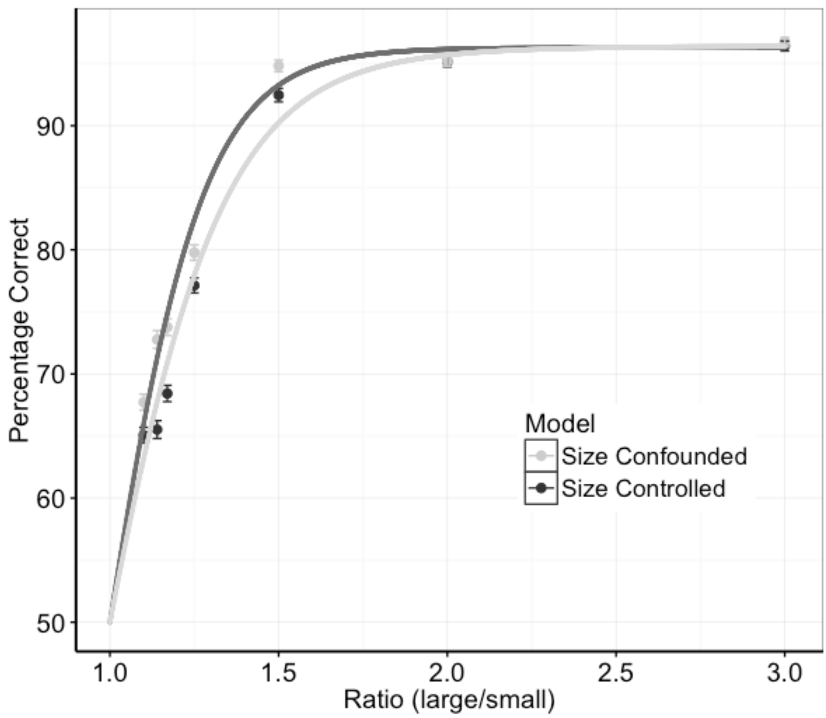


Figure 4.3. *Accuracy on the Quantity Discrimination task as a function of the ratio between number of items of each set. The x-axis shows the seven different ratios presented during the training (1.1, 1.14, 1.17, 1.25, 1.5, 2, 3). The two ways of manipulating surface-area are shown, each with a*

corresponding Weber curve, consistent with students making their choice based primarily on number.

Collapsing all trials disregarding what type of size control they implemented, participants responded correctly on 81.5% of the trials (SD = 8%). We can observe differences in the ratio-dependent performance curve according to the grade the students were attending (Figure 4.4). In general, the older the students were, the better they performed in the Quantity Discrimination task, with the exception that 10th grade did worse than 9th.

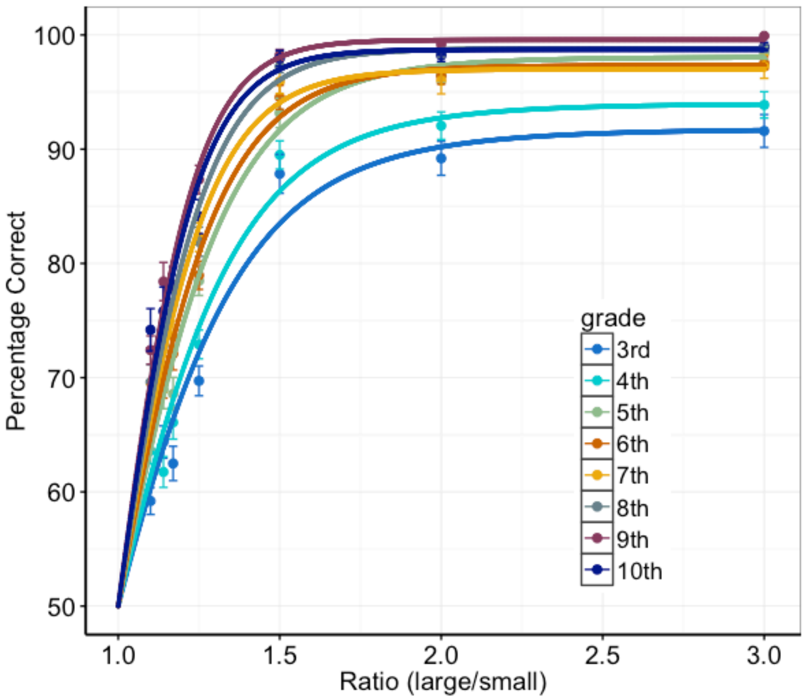


Figure 4.4. *Accuracy on the Quantity Discrimination task as a function of the ratio between number of items of each set. All trials collapsed regardless of their type of control size, and splitted by grades. The x-axis shows the seven different ratios presented during the training (1.1, 1.14, 1.17, 1.25, 1.5, 2, 3).*

4.3.1.2 Response Time

On average, the Response Time was 885 ms (SD = 281 ms). The response time (RT) decreased across successive runs of the Quantity Discrimination task. This indicates that students augmented their speed during the session. In Figure 4.5 we show the mean RT for the group at each run (±SE). A logarithmic slope was computed for each student. Participants had significantly negative training slopes for response time across the session: $t\ (502) = -22.2$, $p < 0.001$. The decreasing response time across successive runs could be an indication that participants improved their performance during the session.

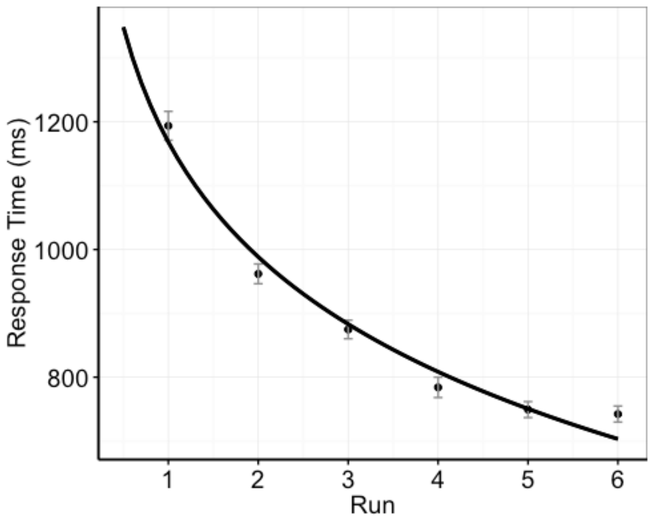


Figure 4.5. *Response Time (RT) on the Quantity Discrimination task as a function of progress during the session (composed of 6 runs). Error bars present ±SE for the group performance. The logarithmic regression line represents the average slope and intercept across participants (n =503).*

4.3.1.3 Efficiency

The efficiency, operationalized as the percentage of correct answers divided by RT, increased rapidly across the first four runs, and maintained the same level for the following two runs (5 and 6), suggesting that there might be a ceiling in participants' performance

(Figure 4.6). It should be noted that, in Quantity Discrimination task, each run had the same structure (35 trials with an increment of difficulty every five trials). Thus, difficulty increases within runs, but not between runs. A linear training slope was computed for each child. Students had significantly positive slopes for efficiency, consistent with their performance improving through the session (t (502) = 19.3, $p < 0.001$; Figure 4.6).

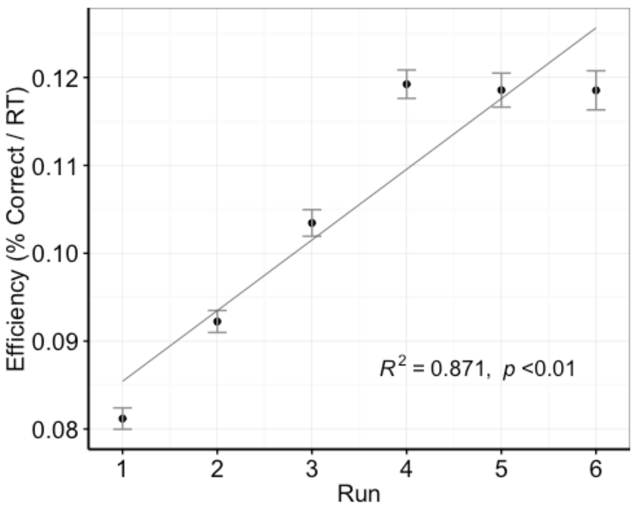


Figure 4.6. *Efficiency (Percentage correct / RT) on the Quantity Discrimination task. Error bars represent ±SE for the group efficiency at each run. The linear regression line represents the average slope and intercept across participants (n = 503).*

4.3.2 Numerical Estimation Task

We recall that in this task students (n= 524) viewed a quantity of briefly flashed items and had to choose the correct match among three possible digits. Students completed a total of six runs, split into two rounds of three runs (round 1= Runs 1 to 3; round 2= Runs 4 to 6). The structure in both rounds was the same, with the three runs increasing in difficult from *Easy* (*Span 6, Run 1 and 4*), to *Medium* (*Span 4, Run 2 and 5*), and *Hard* (*Span 2, Run 3 and 6*). In each of the three types of runs, participants were asked to estimate all set sizes, from 1 to 21 items. Therefore, their performance was affected by **size effect** (size of the collections incremented in each run), and by **distance effect** (span or distance decremented in each run).

4.3.2.1 Percentage of correct answers. Size and distance effects.

On average, students responded correctly on 63.15% of the trials (SD = 7.9%; chance = 33%).

a) Size effect

Consistent with the predictions, students had a higher percentage of correct answers when the size of the collection was smaller. A precise subitizing response for items 1 to 3 can be observed. Starting from size 4, one can notice the difference in participants perception of small and large numerosities (Feigenson et al., 2004; Figure 4.7). However, for older ages from 10 to 16 year olds (5^{th} to 9^{th} grade), size 4 could be considered inside the range of subitizing (Figures 4.8, 4.11). The higher the number of items was, the worse the performance was. Interestingly, the only size collection in which participants performed below chance was the largest one, 21 items (Figure 4.7). Presumably, out of the range of our test, performance for sizes above 21 would not recover from chance response level.

Disregarding the special case of the subitizing range (which does not affect our overall analysis), a simple regression line predicts the percentage of correct answers (PC) as a function of the number of items (N): PC = 104 - 3.6 × N, explaining more than the 90% of the variance in the percentage of correct answers ($R^2 = 0.93$, $p < 0.001$; Figure 4.7).

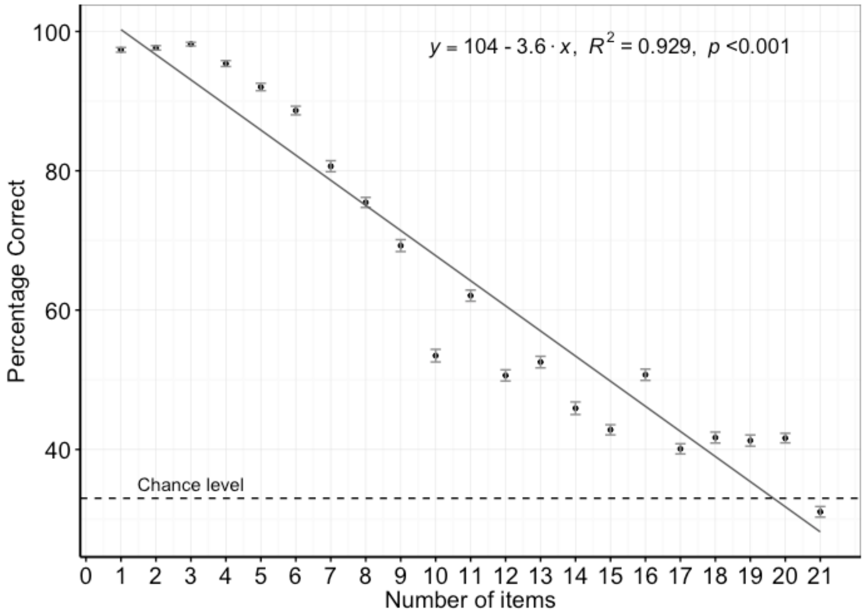


Figure 4.7 *Percentage of correct answers (chance = 33%) on the Numerical Estimation task as a function of the number of items presented in each trial (size effect). Error bars represent ±SE for the group performance. The regression line (P correct = 104 - 3.6 × N; being N the number of items) represents the average slope and intercept across participants (n= 524; R^2=0.93, p < 0.001).*

This model was valid at each school grade. We calculated a regression line (Figure 4.8): PC_3 = 95 - 3.4 x N; PC_4 = 99 - 3.5 x N; PC_5 = 104 - 3.7 x N; PC_6 = 106 - 3.6 x N; PC_7 = 108 - 3.7 x N; PC_8 = 109 - 3.7 x N; PC_9 = 110 - 3.5 x N; PC_{10} = 109 - 3.6 x N. Indeed, the variations in the percentage of correct answers as a function of the number of items (the slope of the regression line), showed very few variations between grades (*M* = -3.6, *SD* = 0.11, *SE* = 0.04). However, the intercepts of the eight regression lines (*M* = 105, *SD* = 5.4, *SE* = 1.9) showed that different grades have different levels of knowledge of the cardinalities 1-21: The higher the grade which the students were attending was, the better the mapping between digits and quantities was. We only found one exception at the 10th grade, whose percentage of correct answers was situated below that of the

9th grade (Figures 4.8, 4.9, 4.10), similarly to what happened in case of other test, the Quantity Discrimination task. We believe that this exception may be due to random characteristics of one of the two grades (delayed 10th grade or advanced 9th grade) and/or to the size of the samples (9th grade, n = 40; 10th grade, n = 39), which are smaller respect to the other grades.

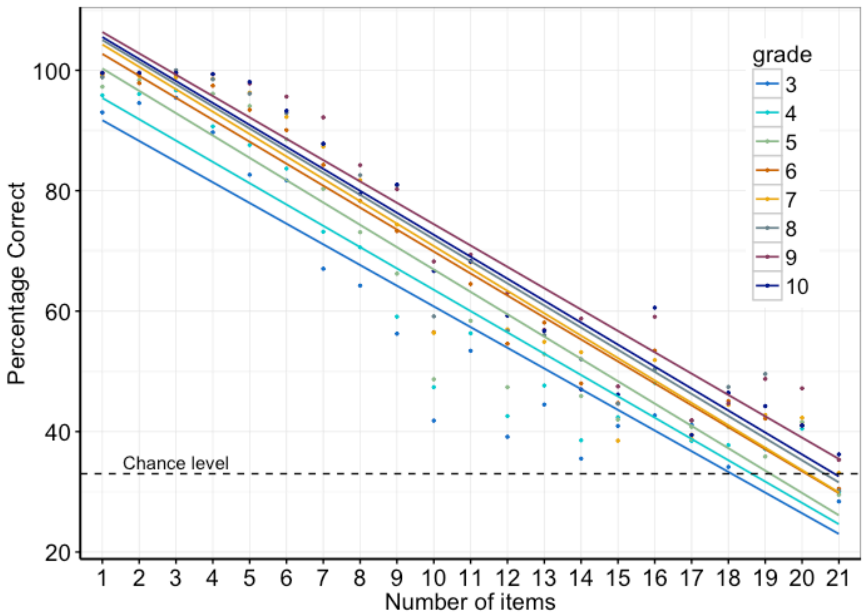


Figure 4.8. *Percentage of correct answers (chance = 33%) on the Numerical Estimation task as a function of the number of items presented in each trial (size effect), and split by grades.*

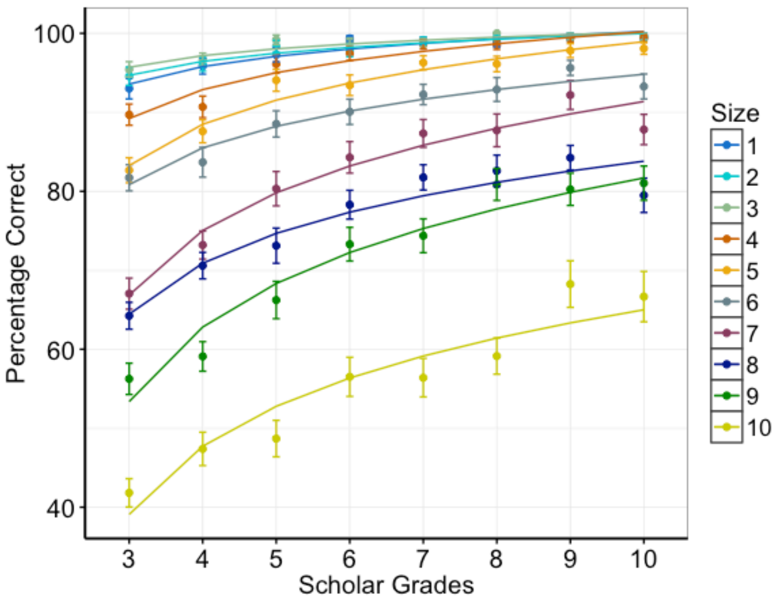


Figure 4.9. *Percentage of correct answers on the Numerical Estimation task for the collections of 1 to 10 items (size effect), and split by grades.*

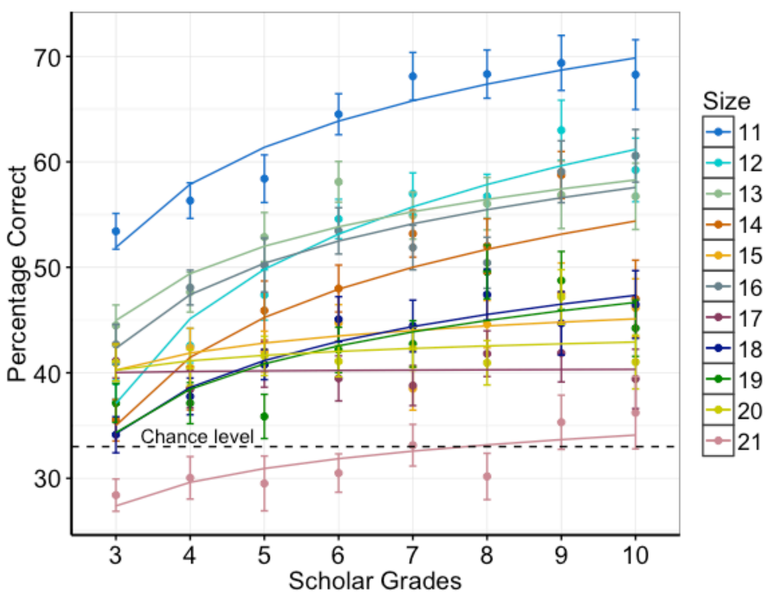


Figure 4.10. *Percentage of correct answers on the Numerical Estimation task for the collections of 11 to 21 items (size effect), and split by grades.*

Performance in lower grades, and especially third (8-9 year olds) and fourth (9-10 year olds) grades, dropped more markedly when the collections were formed by four or more items. This could indicate an age-dependent distinction between small and large numerosities that was still being developed at these ages. It has been reported this development during the first year of life (Coubart et al., 2014; Figure 4.11) .

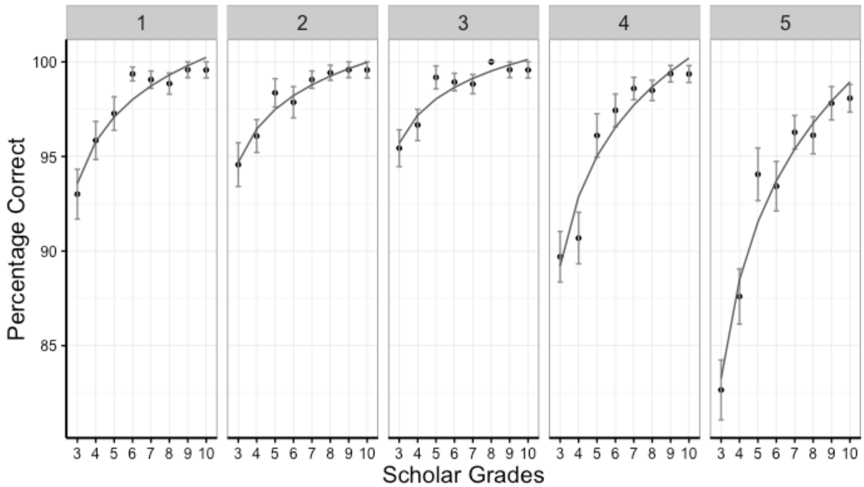


Figure 4.11. *Percentage of correct answers on the Numerical Estimation task for the collections of 1 to 5 items (size effect), split by grades.*

b) Distance effect

Participants' accuracy across the six runs is displayed in Figure 4.12. Consistent with predictions, participants did better on the Easy and Medium runs, in which the distances between digits were bigger (*Span 6* and *4* respectively), than in Hard runs. Repeated measures ANOVA revealed a main effect of Difficulty ($F_{(2,1046)}$ = 1439, p < 0.001). Students were above chance at all levels of difficulty as revealed by planned t-tests: Easy, t (523) = 87.55, p < 0.001; Medium, t (523) = 70.73, p < 0.001; Hard, t (523) = 59.93, p < 0.001.

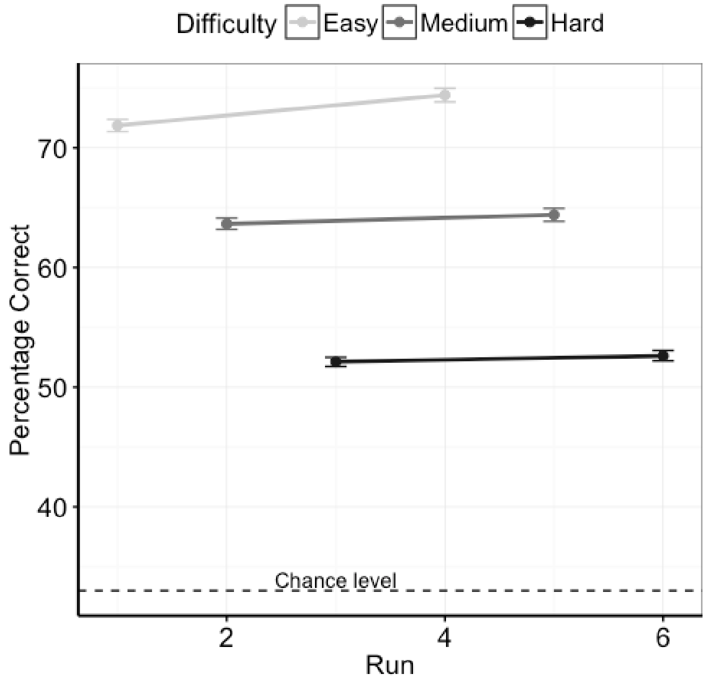


Figure 4.12. *Percent Correct (chance = 33%) on the Numerical Estimation task across 6 runs in one session. Error bars represent ±SE for the group performance. The three linear regression lines (one for each level of Difficulty (Easy, Medium, Hard) represent the average slope and intercept across participants (n = 524).*

Participants' correct responses improved from the Easy run of the first round to the Easy run of the second round (that is, Run1 and Run4; M_{Run1}=71.8%, SE_{Run1}=5.1%; M_{Run4}=74.4%, SE_{Run4}=5.6%; t (1012) = -3.32, p < 0.001). The improvement of the Medium and of the Hard runs of the two rounds (respectively, (Run2 and Run5, Run3 and Run6) was not significant.

The distance effect showed to be consistent with each small variation of the numerical distance between the three digits proposed for selection in active trials. In Figure 4.13 each level of difficulty (numerical distance) is splitted in two: Easy (Spans 6 and 5), Medium (Spans 4 and 3), Hard (Spans 2 and 1); being numerical distances for Span 6: (0, 3, 6) and (6, -3, 0); for Span 5: (-3, 0, 3);

for Span 4: (0, 2, 4) and (-4, -2, 0); for Span 3: (-2, 0, 2); for Span 2: (0, 1, 2) and (-2, -1, 0); and finally, numerical distances for Span 1: (-1, 0, 1). The percentage of correct answers decreased as the numerical distance indicated by the spans decreased gradually.

.3.

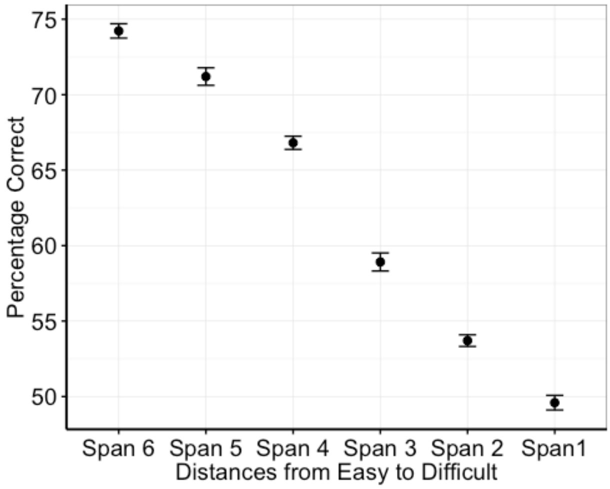


Figure 4.13. *Percent Correct (chance = 33%) on the Numerical Estimation task by type of run difficulty (n = 524). Error bars represent ±SE for the group performance. Each level of Difficulty Easy, Medium, and Hard is now split in two: Easy (Spans 6 and 5), Medium (Spans 4 and 3), Hard (Spans 2 and 1); being distances for Span 6: (0, 3, 6) and (6, -3, 0); distance for Span 5 (-3, 0, 3); distances for Span 4: (0, 2, 4), (-4, -2, 0); distance for Span 3 (-2, 0, 2); distances for Span 2: (0, 1, 2), (-2, -1, 0); and distance for Span 1 (-1, 0, 1).*

c) Size and Distance effects combined

The distance effect is also accompanied by a size effect. For equal numerical distances, performance decreased as the numbers to be compared became larger (Figure 4.14).

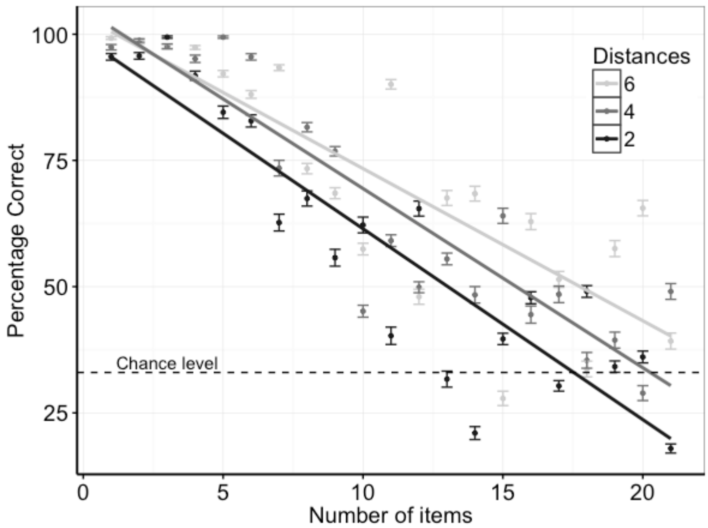


Figure 4.14. *Percentage correct as a function of the number of items presented in each active trial (size effect). Regression lines are separated by distance (distance effect). The horizontal dashed line represents random performance.*

d) Position of the target

The position of the target between the three digits was another factor that affected student performance. When the target was positioned in the middle, the percentage of correct answers reached a 74.8% (chance = 33%), a high value in all grades (*SD*= 0.7). However, when the position of the target in the triplet was on the left (*M*= 60.7%, *SD*= 8%) or on the right ((*M*= 57.8%, *SD*= 6.3%), the percentage of correct responses decreased, especially in lower grades (3rd, 4th and 5th; Figure 4.15).

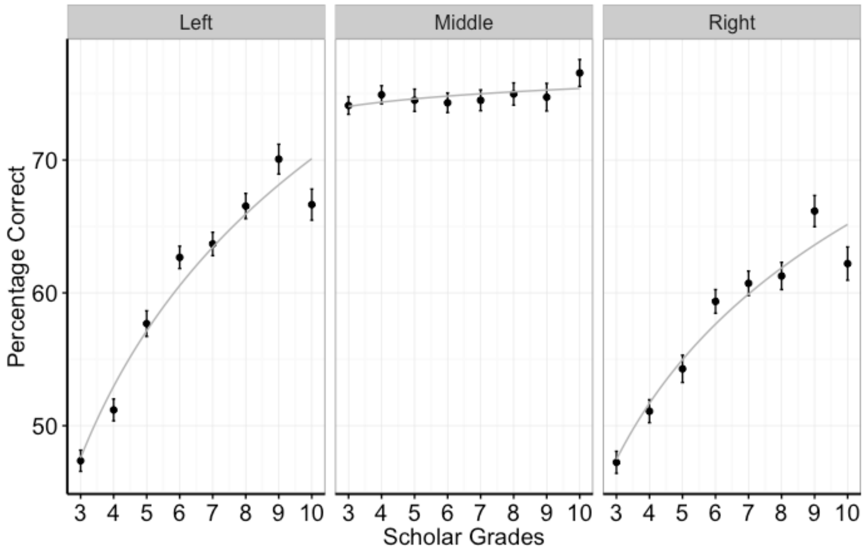


Figure 4.15. *Percentage of correct answers in the Numerical Estimation task split by position of the target number (left, middle, right) and by school grade.*

This position effect could also be observed in the three difficult levels of the Runs (Easy, Medium, Hard; Figure 4.16). On average, considering all participants together (n=524), students responded correctly on 63.15% of trials (SD = 7.9). When the target was in the middle of the triplet, the responses in all school grades, and for all levels of difficulty, were above average (Figure 4.16). This was not the case when the target was on the side of a triplet: a) for hard difficulty trials, all grades had percentage of correct answers below the average but above the chance level, regardless of the side of the target (left or right; Figure 4.16); b) for medium difficulty trials, only 3rd, 4th and 5th grades had a value below the average, regardless of the side of the target (left or right; Figure 4.16); c) for easy trials, 3rd grade had a value below the average, regardless the side of the target; and regarding 4th and 5th grades when the target was at the right, their values of percentage of correctness were also below the average (Figure 4.16).

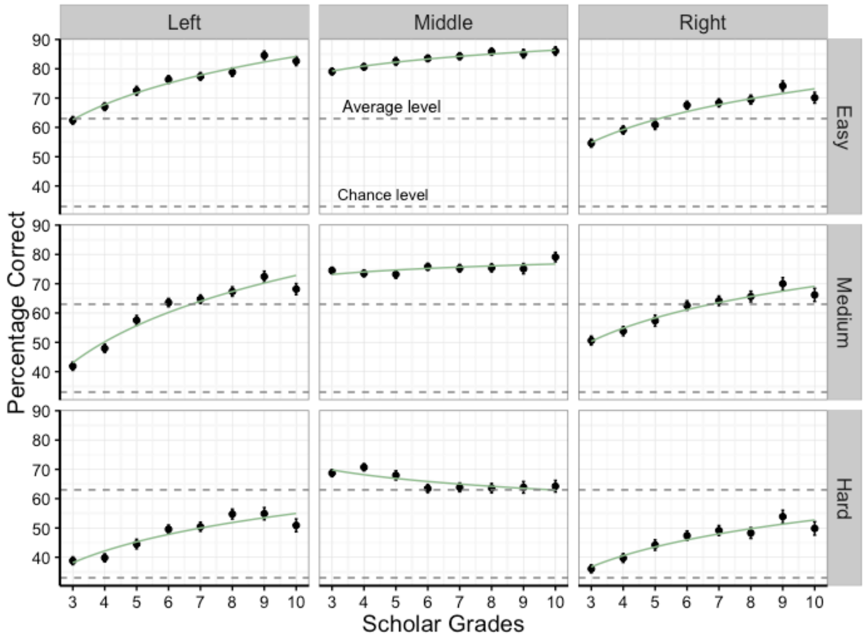


Figure 4.16. *Percentage of correct answers on the Numerical Estimation task split by position of the target number (left, middle, right), the difficulty of the trials (Easy, Medium, Hard) and by the scholar grades that participated in the study. The horizontal dashed line at 33% represents random performance. The horizontal dashed line at 63% represents the average performance.*

We speculate that the level of executive functions required to inhibit the election of the easiest position (middle) in order to choose the correct answer when it is in one side of the triplet, could not be developed until 6th grade (12-13 year olds).

4.3.2.2 Response Time

We analyzed student's response time (RT), that is, the time it took participants to make their choices on each trial. On average, the Response Time was 2040.23 ms (*SD* = 643.08 ms). Figure 4.10 shows the mean RT for the whole sample for each run (±SE). A logarithmic slope was computed for each child through the session. Students had significantly negative slopes for response time through the session: t (523) = -7.25, $p < 0.001$; Figure 4.17. The decreasing response time across successive runs is an indication of improving performance during the session.

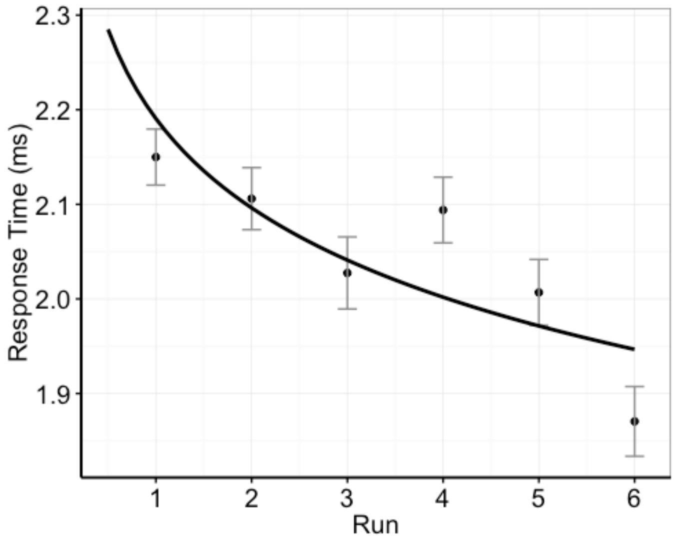


Figure 4.17. *Response Time (RT) on the Numerical Estimation task through the session. Error bars represent ±SE for the group performance. The logarithmic regression line represents the average slope and intercept across participants (n = 524).*

4.3.2.3 Efficiency

The efficiency, operationalized as the percentage of correct answers divided by RT, decreased within each round (round 1 = Run 1, 2, and 3; round 2 = Run 4, 5, and 6) due to the increasing level of difficulty. However, between runs of equal difficulty, the efficiency increased significantly (easy runs: t (1004) = -4.17, $p <$ 0.001; medium runs: t (1007) = -3.5; hard runs: t (999) = -4.67, $p <$ 0.001, $p < 0.001$; Figure 4.18), showing that students progressed through the session, and suggesting that their ability to map digits to quantities, ie their knowledge of cardinality, could improve with practice.

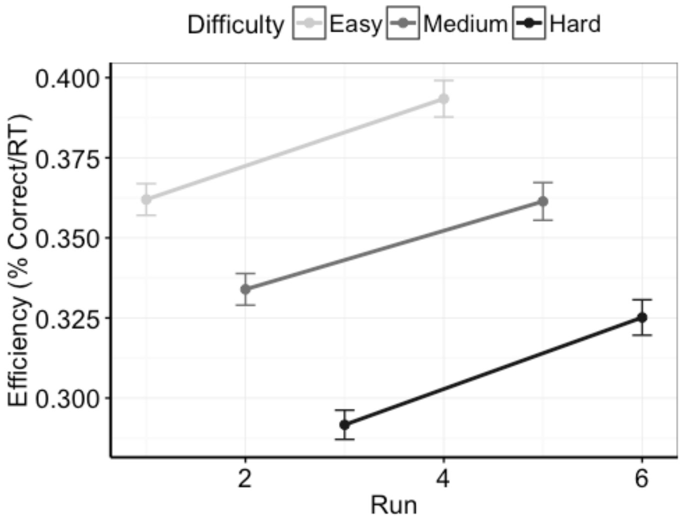


Figure 4.18. *Efficiency (Percentage correct / RT) on the Numerical Estimation task. Error bars represent ±SE for the group efficiency at each run. The three linear regression lines (one for each level of Difficulty (Easy, Medium, Hard) represent the average slope and intercept across participants (n = 524).*

4.3.3 Relation between the numerical cognition tasks and between school math performance

4.3.3.1 School Math Performance

We considered students' math marks provided by the school the measure of students' school math performance. The school year was divided into three quarters that conclude in three evaluations: from the beginning of the course in mid-September until the Christmas holidays takes place the first quarter. The second quarter begins in January and ends in March approximately, depending on the date of Easter holidays; and finally, the third and last quarter of the scholar year begins in April and ends in June, approximately. Each quarter has its own qualification: first, second, and third evaluation.

Considering that the twenty-four classes involved in this research took their one-hour session test during April and first week of May, the situation on the student's math performance at the time of the investigation could be appropriately determined by the second and third evaluations, i.e. by school marks obtained in the quarters before and after our session test.

School marks range from 0 to 10. Any rating below 5 indicates that the quarter was not been successfully passed. Below, we analyze the correlation between the school math marks and the performance in the numerical estimation and the quantity discrimination tasks. In Appendix 1, we present the analysis of the assumption of normal distribution of school math marks. Because the math marks did not follow a normal distribution in any of the grades that participated in the study, to analyze the correlation between them and performance in the numerical estimation and quantity discrimination tasks, we used non-parametric tests.

4.3.3.2 Correlation between school math performance and numerical cognition tasks: Numerical Estimation (Digits game) and Quantity Discrimination (Panamath game)

To analyze the correlation between students' school math marks and their performance at numerical cognition tasks (Numerical Estimation (NE), and Quantity Discrimination (QD)), we used a non-parametric statistic: Spearman's correlation coefficient (r_s), because the school marks were not normally distributed (see Appendix 1). We tested correlations between school math marks and two main measures of our tasks, namely, percentage of Correct answers (PC) and Response Time (RT). We studied the correlation of these measures and the school math marks at 2nd and 3rd Evaluation (Tables 1 to 8). We consider that using both school math marks instead of only one, is a conservative procedure that provides us more stability to avoid some transient and unrepresentative fluctuations of some students.

Each grade may differ from those of previous years, with different group dynamics and level of performance, also with respect to other grades of the same year. School marks are adapted to the idiosyncrasy of each grade for each year. That is why, mixing all the grades as if the marks were consistent among all of them would be a mistake and introduce noise in the measures. We have chosen to carry out the analysis of each grade separately, in a more laborious process but that we consider more correct.

Appendix 2 reports the detailed analyses of the correlations on each age group. Below, we will report the summary of these analyses.

First, because not all students of 3rd, 4th and 6th grades could perform both tasks, to analyze correlations between tasks and school math marks, we considered the smallest sample of both tasks for that specific grades. Figure 4.18a shows the number of participants by grade that performed both tasks, and that were used to calculate the correlations. Should be considered in the interpretation of the results that 9th and 10th grades represented the smallest samples.

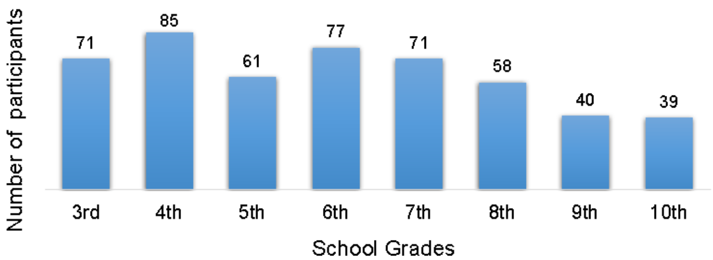


Figure 4.18a. *Number of participants by grade that performed both tasks, Numerical Estimation task and Quantity Discrimination task.*

Figure 4.18b shows a summary of the number of evaluations (none, only one: 2nd or 3rd, both: 2nd and 3rd) in which math marks correlate with the percentage of correct answers in Numerical Estimation task (NE) and Quantity Discrimination task (QD). When percentage correct correlates only in one evaluation with math marks, we consider this correlation unrepresentative to extrapolate to general school math performance of the student. This was the case of QD task for 4th grade where the correlation was significant only at 3rd evaluation, and for 8th grade where the correlation was significant only at 2nd evaluation (Figure 4.18b).

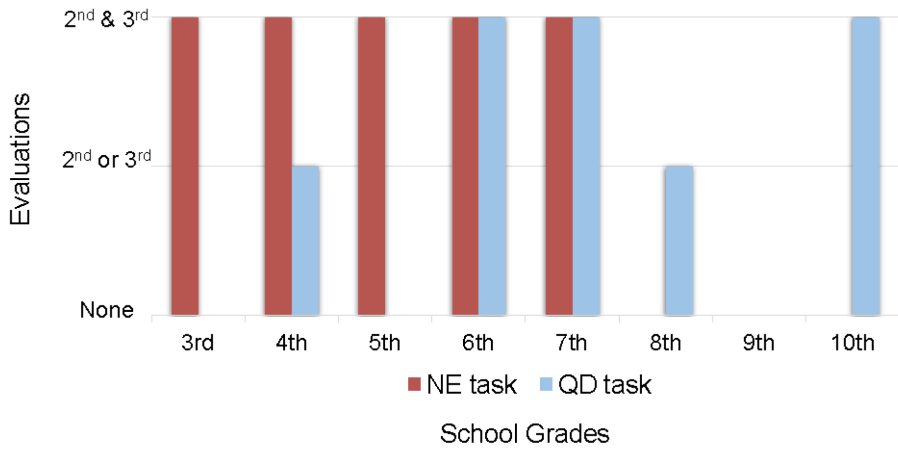


Figure 4.18b. *Number of evaluations (none, only one: 2nd or 3rd, both: 2nd and 3rd) in which math marks correlate with the percentage of correct answers in Numerical Estimation task (NE) and Quantity Discrimination task (QD) splitted by school grades.*

A brief mention about the correlation between the percentage of correct answers in QD task and school math marks at 10th grade. While this result is interpretable since the correlation is significant in both evaluations and also goes in the direction of previous literature (Libertus et al., 2011), we want to be cautious on its interpretation because of three reasons: 1) at this age the cohort was quite reduced with respect to the previous age groups, 2) the other grades with the same pattern (correlation significant at both evaluations) in QD task were only 6^{th} and 7^{th} grade. Thus, we can not explain the result in terms of a range of ages, but rather it seems an isolated result. 3) This grade showed an unexpected pattern in both tasks because its percentage of correct answers was situated below that of the 9^{th} grade (Figure 4.4 for QD task, and Figures 4.8, 4.9, 4.10 for NE task). This may be due to random characteristics that could confer a special idiosyncrasy to the sample of 10^{th} grade.

Considering the numerical cognition tasks that we are analyzing, the percentage of correct answers alone is a variable that better explains the performance on the task than the response time alone. Having said this, there are two grades, 6^{th} and 10^{th}, in which the response time of QD task positively correlated with school math marks, indicating that pupils who responded more slowly also had higher math marks.

4.3.3.3 Correlation between numerical cognition tasks: Numerical Estimation (Digits game) and Quantity Discrimination (Panamath game)

The performance in both task, NE task and QD task, is described by the percentage of correct answers mainly, and the response time secondarily. We wanted to know if both measures separately correlated between tasks.

Appendix 2 reports the detailed analyses of the correlations on each age group. Below, we will report the summary of these analyses.

The percentage of correct answers correlated between tasks from 4^{th} to 8^{th} grade. Considering that percentage correct is the main measure that explains the performance in the tasks, overall the data suggests that the tasks could be related at this range of grades (4^{th} to 8^{th}). The response time correlated also between tasks from 3^{rd} to 8^{th} grade, with the only exception of 6^{th} grade. Thus, the speed of responses had similar profiles in both tasks.

Unlike other grades, in 9th and 10th grades no significant correlation was observed in performance between tasks. These two grades corresponded to the older students, from 14 to 16 year olds.

4.3.3.4 Correlation between Percentage of Correct answers and Response Time

Appendix 2 reports the detailed analyses of the correlations on each age group. Below, we will report the summary of these analyses.

Considering both tasks separately, the correlation between the percentage of correct answers and the response time was significant and positive in both tasks for all grades, with the only exception in 9th grade where the correlation was significant only in QD task.

Overall, the longer the time to respond was, the higher the percentage of correct answers was. Therefore, a speed-accuracy tradeoff clearly appeared. This fact may be the basis for specific pedagogical guidelines, suggesting that being quicker at responding should not be encouraged, at least in numerical cognition tasks, since it's no virtue to go fast while making correct performance worse.

4.4 Discussion

Our Studies 1 and 2 showed that a quantity discrimination training and a numerical estimation training have different effects on symbolic arithmetic. While training the ANS through a quantity discrimination task transferred improvements in symbolic math only when exact calculations were not required, training the Digit-Quantity mapping through a numerical estimation task transferred improvements to symbolic math both when an exact answer was required and when it was not. The results of Study 1 and 2 are consistent with the fact that the Approximate Number System is not sufficient to compute exact operations with numbers, because these calculations also require the establishment and understanding of the exactness of mathematical language (Bonny & Lourenco, 2013; Butterworth, 2010; Dehaene, 2001; Gordon, 2004; Lemer et al., 2003; McCrink et al., 2013; Pica et al., 2004). Since Arabic digits are the precise representation of quantities, training this mapping seems to facilitate calculations with numbers, thus percolating into an improvement of arithmetic skills (Figure 9, Study 2). Ours is not the only study that shows the importance of a proper understanding of the exactness of mathematical language; Mundy & Gilmore (2009) also found that the accuracy in mapping Arabic digits and nonsymbolic representations is related to mathematics achievement in 6-8-year-old children.

ANS and symbolic math abilities relay on different cognitive mechanisms: left inferior parietal cortex is specialized for symbolic numbers processing, while right superior parietal lobule for nonsymbolic quantities (Dehaene & Cohen, 1991, 1997; Sokolowski et al., 2017). Furthermore, to improve the accuracy in mapping Arabic digits to quantities, learning could have a more prominent role than the improvement of ANS accuracy, which rely on very primitive abilities shared by all humans and many nonhuman animals. In this study, we explored both abilities in a school environment, testing a wide range of ages going from 8 years to 16 years of age.

Our first task has been to assess the kinds of phenomena that characterize the two tasks. Overall, we showed that the quantity discrimination test exhibited the signature of the ANS. This is a ratio-dependent performance predicted by Weber’s law that results in a specific curve of percentage of correct answers as a function of ratio (Figure 4.3; Feigenson et al., 2004; Libertus & Brannon, 2009; A. B. Starr, Libertus, & Brannon, 2013). We also found that the performance curves obtained for each grade improved with the age of the students (Figure 4.4). The smaller the “large set/small set” ratio was, the more difficult the task of discrimination was. The ceiling performance (around 90%) for grades 3rd and 4th (8-10 years) was approximately at ratio 2. The ceiling performance (between 97 and 100%) of older students was located at ratios 1.5 to 1.75. These data confirm that ANS accuracy increases with age, at least from 8 to 16 years (Halberda & Feigenson, 2008).

Also, the Number Estimation task was characterized by a set of known signatures such as size and distance effects. Indeed, the percentage of correct answers was exhibited the size, the distance, and the "position of the target" effects. The size effect appears in the different rates of correct responses as a function of set size. Participants gave more correct answers to smaller sets sizes (Figure 4.7). For collections of one, two, or three items (and four items for older students, Figures 4.8 and 4.11), performance was almost flawless. This is the range of subitizing (Revkin et al., 2008). Beyond this range, performance deteriorates, to the point of getting close to random for high numerosities. Notably, participants were practically always above chance, with the exception of the biggest set size (21; Figure 4.7), and even in that case, at least the older students performed above chance (Figure 4.10). Thus, we could verify that accuracy in the mapping between Arabic digits and quantities was affected by the size of the collections and improved with age, from 8 to 16 year olds (Figures 4.8, 4.9, 4.10, 4.11). Regarding the distance effect, participants performed better in the runs with bigger distances between digits (Figures 4.12 and 4.13). The distance effect was also accompanied by a size effect. For equal numerical distances, performance decreases as the numbers to be compared become larger (Figure 4.14). Finally, the position of the target (left, middle, right) influenced the correctness of the responses. When the target was in the middle, the percentage of correct answers was definitely higher than when it was on the sides (middle: 75%, left: 60%, and right 58%). This effect was more marked for the younger children (3rd, 4th, and 5th grades; Figure 4.15). Also, the effect of the position affected especially the most difficult runs (smaller distance between digits), where all grades performed below the average when the target was on one side, but above the average when the target was in the middle (Figure 4.16). Thus, position effect could be age and difficulty affected. We speculate that executive functions are required to inhibit the election of the easiest position (middle) in order to choose the correct answer when it is presented on one side of the triplet.

We have also analyzed the relation between the percentage of correct answers and the time to respond, in both tasks separately. Results indicate that slower response time gave more correct answers in both numerical cognition tasks. We have two suggestions; first, it should be reconsidered the interpretation of the response time in numerical cognition tests since the longer the time to respond was, the higher the percentage of correct answers was; and second, in numerical cognition tasks, being quicker at responding should not be encouraged by the teachers at school or by experimenters at laboratories.

But our main questions concern the relations between the two tasks (and whether this relation changes across ages) and the power that these tasks have to predict school math performance (and whether this too changes across ages). As for the first point, in most of our cohorts (from 4^{th} to 8^{th} grade) the percentage of correct answers in the two tasks positively correlated. The exceptions were the younger cohort (3^{rd} grade), and the two oldest cohorts (9^{th}, and 10^{th} grades). The response time, however, correlated between tasks from 3^{rd} to 8^{th} grade, with the only exception of 6^{th} grade. Thus, the speed of responses had similar profiles in both tasks.

The interpretation of the lack of correlation between the task in the older age groups is difficult, because the sample size is smaller than the other samples. Lack of sufficient power may have hidden the existence of relations among the tasks. Also, it has to be considered that, precisely only these two grades, had an unexpected pattern in both tasks since grade 10^{th} had a percentage of correct answers situated below that of the 9^{th} grade (Figure 4.4 for QD task, and Figures 4.8, 4.9, 4.10 for NE task), which may reflect some random characteristic of one or both of the samples of these two grades. In the case for our youngest participants, we suggest that the dissociation between the two tasks in percentage of correct answers is real. Remember that this age is also the closest one to the participants in Studies 1 and 2, where we found differential effects of prolonged trainings with a Numerical Estimation task and a Quantity Discrimination task. A possible explanation of these results could be as follows. At a young age, when knowledge of the meaning of the digits is limited, the digit-quantity mapping ability is relatively independent from ANS. The link between ANS and the symbolic mathematical language is still to be properly established. It is also at this age training this ability may have its strongest effects, precisely because the understanding of the exactness of the relation between digits and quantities does not fall out automatically from an already existing ANS. With learning, and the improvement of the notion of cardinality, the bridge between ANS and exact determination of the quantities may close, and hence a better ANS can, not only provide approximate evaluations of quantities, but also improve the efficiency of the digit-quantity mapping.

Our second main question focuses on the relation between school math performance and the two tests we studied. Overall, the Quantity Discrimination task correlated with school math grades at grades 6th, 7th and 10th (or at 12-13 and 16 years of age), where “clearly” here means that the correlation held for both school evaluations. It is difficult to interpret a result not showing a uniform pattern across different ages. For 6th and 7th grades, correlation was positive, suggesting that in some limited cases ANS can indeed predict school performance. However, the case of the isolate result for 10th grade, could respond to other aspects such as the size of the sample (the smallest one), or some random characteristics that could confer a special idiosyncrasy to the students of this grade. It should be recalled that precisely 10th showed an unexpected pattern in both tasks because its percentage of correct answers was situated below that of the 9th grade (Figure 4.4 for QD task, and Figures 4.8, 4.9, 4.10 for NE task).

Importantly, clearer results appeared when inspecting the Numerical Estimation task. There, we found a remarkable correlation between the school math marks and the results in that task. This occurred for an extended period in development going from 3rd to 7th grade, or from 8 to 13 years of age, that may establish the foundations for better learning and confidence in later years.

Although a correlation does not imply the existence of a causal relation, nor does it allow us to draw conclusions about the direction of this hypothetical causal relation, we suggest that it is a better knowledge of the meaning of the digits that can cause children to be better in math at school. We have shown in Study 2 that training the Digit-Quantity relation translates into a generalized improvement in mathematical abilities, and not only in one particular aspect of mathematical knowledge; as well as other studies also indicated it (Booth & Siegler, 2008; Mundy & Gilmore, 2009).

The interpretation of the lack of this correlation in older age groups (from 14 to 16 years of age) is difficult. Besides a problem of sample sizes, since precisely the last three grades had the smallest samples (Figure 4.18a), specially 9th and 10th grades; we suggest here other explanations. A possible explanation is that their knowledge of cardinality, at least until 21 (the biggest set of items in our test), is sufficient to solve the requirements from school mathematics. In addition, the school curriculum for older students, with an increase of complexity and abstraction, for example in terms of algebra, trigonometry, statistics…, could be less dependent of the knowledge of cardinality of the first 21 Arabic digits. And finally, a conceptual reason could be that in order to get to a correct answer in a mathematical operation, the understanding that the solution is precise and unique is fundamental. Answering that 3+2 is equal to 6 gives a good approximation of the real result of the operation, but does not count as a correct answer. A correct answer, and with it the awareness of the existence of a mistake, can only come once one realizes that “5” means exactly five. Possibly, again, the consolidation of this concept is fundamental in the very first years of mathematical practice, and once acquired, does not have the same role. Thus, older students may not benefit of a finer ability to perform a numerical estimation task because they have already acquired and interiorized the awareness of the crucial importance of exactness in order to understand cardinals.

The proposed interpretations have to be further explored, but we believe that they can provide the blueprint for further educational interventions, which we hope to be able to realize in the coming years.

5. GENERAL DISCUSSION

5.1 The role of ANS training on symbolic mathematics performance

In this section, we want to summarize the main results of our work, and propose some speculative hypotheses of why we found what we found. We insist on the word “speculative”, hoping that the reader understands that the aim of our dissertation was less theoretical than practical: finding ways to improve children’s mathematical performance in particularly crucial school years. Theory helped us to find practical ways to address our question. Now the result of our practices could be helped to give back something to what the theory gave us: some insights in what may have happened and why in our training studies. But here is where our dissertation stops and another, deeper and longer work, should begin. We can only point at the directions that this work should take.

In Study 1 (chapter 2; Ferres-Forga, Bonatti, et al., 2017) and Study 2 (chapter 3; (Ferres-Forga, Halberda, et al., 2017) we measured the arithmetic competence of ninety-one, and ninety, 7-8 year old children respectively, with two six-minute paper-and-pencil tests of symbolic math. One test for additions and one for subtractions were presented in the typical formats used in the schools (Figures 1a and 1b; Studies 1 and 2). We also measured pupil’s initial level of mathematical reasoning with a six-minute test of operation-deduction problems (3 ☐ 5 = 15; Figure 1c; Studies 1 and 2). In it, participants had to find the operation (addition, subtraction or multiplication) that rendered the equality true. In both studies, a group of participants who trained the ANS (Approximate Number System) with a Quantity Discrimination task by playing the Panamath program.

We first verified that our ANS tests related with formal mathematic abilities, as expected by previous literature (Feigenson et al., 2013; Halberda et al., 2008; Libertus et al., 2012), showing that individual differences in ANS efficiency were related to the number of correct answers in the math pre-test (Figure 5, Study 1). We also verified that our ANS training showed the typical ANS signature: a ratio-dependent performance predicted by Weber's law. This resulted in a specific curve of percentage of correct answers as a function of ratio (Figure 2, Study 1; Figure 3, Study 2; Dehaene, 1997; Feigenson et al., 2004; Libertus & Brannon, 2009; A. B. Starr et al., 2013). Finally, we checked that precision in the ANS could be trained (Knoll et al., 2016; Obersteiner et al., 2013; Park et al., 2016; Park & Brannon, 2014): indeed, ANS efficiency increased over the course of the 3-week training period (Figure 3, Study 1; Figure 5, Study 2) and even across the three runs within each day (Figure 4, Study 1).

We then looked at the effects of training on children's mathematical performance. Training the ANS transferred to improvements in post-training symbolic math tests. However, in both studies, benefits were mainly apparent in the operations test. Specifically, in the first study only those children with a low-profile in arithmetics improved, and only in the operations test; in the second study the improvement was not limited to low-performing children.

Let us focus on what improvement in our operation test may indicate (Figures 11-12, Study 1; Figure 9, Study 2). The ability to solve a problem presented in a format such as: 3 ☐ 5 = 15, requires a comprehension on how the result varies as a function of the operation to be performed on the operands, that is, on how each arithmetic operation changes quantities. Stating it in a very simple way, the result is "more" for addition, "less" for subtraction, and "much more" for multiplication. This knowledge may be enough to correctly solve the problems, without the need of calculating the exact result of the operation. Even solving the exact calculation wrongly, but approximately right, could still be enough to solve the problem satisfactorily. This is an important aspect of mathematical knowledge, but it is not what makes one be able to solve mathematics operations correctly.

It is indeed true that there was a remarkable aspect of this improvement. During ANS training, the arithmetic rules were neither explained nor practiced. Children knew it beforehand, because at this age they practice additions, subtractions and some multiplications in school. But nevertheless, the fact that comparing two quantities of dots for a 3-week period improved how the way some children understand what an operation does is something to be stressed. Possibly, the training modified how those children who did not understand how the arithmetic operations change the quantities, but only applied the calculation algorithms mechanically, see what an operation does. This could be especially true for multiplication. The complexity of this operation is bigger than that of a subtraction, and that of a subtraction is bigger than that of an addition. Children are sensitive to this hierarchy (Knops et al., 2014; Linsen et al., 2014; Prado et al., 2011). However, the fact that in Study 1 low-performing ANS Training group improved on all types of operations in the Operations subtest (Additions, Subtractions, Multiplications; Figure 12, Study 1) suggests that the benefits that they received from the training are generalized to any operation.

Another issue to consider in the resolution of the operation test is that, since it was not strictly necessary to perform the calculation (and in case the child did it, it was not essential that the result was exact, just a good approximation could be enough), it follows that it was not needed, although desirable, a precise mapping between digits and quantities. Thus, just with an approximate mapping the child still had chances to solve correctly the problems presented in the operations tests.

To recap, to solve the problems in the operations test, the understanding of the arithmetic operations was needed, although it was not needed the exact calculation; the basic arithmetic rules ("more" for additions, "less" for subtractions, and "much more" for multiplications) were not explained nor practiced in the ANS training, so they had to be already known since there was a baseline of correct answers in the pre-test; and finally, an approximate mapping between digits and quantities could be enough and, in any case, it was not practiced in the ANS training either. Then, why training the ANS by quantities comparisons transferred into a significant improvement in the operations test in both studies, with different participants?

We suggest that with the ANS training, children may get used to think about quantities, the meaning of the digits, even if they do not know precisely the mapping between particular digits and its particular quantities, – something neither reinforced nor needed in this training. This feeling of quantities can then help in the task of comparing what unknown operation best feet an equation like 3 ☐5 = 15, because it can help them to look for “more”, “less” or “much more” in the result of that operation. After all, there are similarities between the task of comparing the two sets of items at both sides of the screen looking for “where are more?” in the ANS training and the task of comparing at both sides of the equal sign, 3 ☐5 = 15, looking for “more”, “less” or “much more”. The quantity comparison in the training could have helped children to think about magnitudes (the meaning of numbers). We may speculate that this cognitive mechanism could be related with the activation of the intraparietal region of both hemispheres (Izard et al., 2008), where the approximate number system is located.

Consistent with this explanation are two facts: a) that mostly low-performing children benefited from ANS training (that is, those children who may have not assimilated what an operation does); b) that improvement did not transfer to the additions and subtractions tests, in neither of the two studies, because in these tests a precise calculation is required.

At the same time, the results obtained are consistent with the idea that the existence of a number sense (ANS) is not sufficient to explain our ability to compute exact arithmetic operations, to follow calculation algorithms with numbers, and in general, to progress towards more advanced mathematical abilities (Butterworth, 2010). An accurate symbolic mathematical language is required (Bonny & Lourenco, 2013; Dehaene, 2001; Gordon, 2004; Lemer et al., 2003;

McCrink et al., 2013; Pica et al., 2004). The language of Arabic digits offers exactly that: a precise representation of quantities.

Although the potential benefits of our extended ANS training are not as far-reaching as one might have hoped, there may still be a place for this type of training. Our suggestion is that for younger and lower-achieving students this training could be a complement to give them the opportunity of thinking about quantities and to give them a sense of magnitudes in an effortless and playful way.

However, could it not be more beneficial to directly train the mapping between digits and quantities, so that the child gets, not only a sense of, but also the value of these magnitudes, when she thinks about numbers? This hypothesis led us to the main idea of our novel training, realized in Study 2.

5.2 Training the Digit-Quantity mapping and its effects on symbolic mathematics

In Study 2 (chapter 3; Ferres-Forga, Halberda, & Bonatti, 2017) we independently trained the mapping between Arabic digits and quantities with a Numerical Estimation task, by asking chidren to play the Digits program for a three-week period. Other children were trained in ANS precision with a quantity discrimination task, playing the the same Panamath program, for the same time, the same days of the week, in the same class and with the same computers.

We verified that the precision of the mapping between digits and quantities was trainable, since performance efficiency increased during the course of 3-weeks of training (Figure 8, Study 2). This is consistent with the increasing functional specialization of the left parietal cortex due to arithmetic activity and mathematical language development, such as codification of Arabic digits (Ansari & Dhital, 2006; M. Piazza et al., 2007; Manuela Piazza & Izard, 2009; Rivera et al., 2005).

Our main result was that strengthening the mapping between digits and quantities, resulted in a generalized improvement of 7-year-olds' mathematical competence. Training the relation between Arabic digits and their non-symbolic quantity representations enhanced the understanding of the *exact* relation between digits and their meaning. It makes children focus on the cardinality that digits refer to. This is, most likely, what makes the significant improvement in arithmetics skills possible (Figure 9, Study 2). We measuerd arithmetic competence with the same three math tests used in Study 1 (Figures 1a and 1b; Studies 1 and 2). According to this common yardstick, the Numerical Estimation training resulted in a consistent improvement piling on top of any improvement that a regime based on training ANS may have provided (Figure 9; Study 2). These results confirm our hypothesis. Thus, we suggest that training the mapping between digits and quantities may generate substantial benefits, improving mathematical competence in 7-8-year-old children. Our results are consistent with the finding that in 6-8 year olds the accuracy of mapping Arabic digits to nonsymbolic representations is related to mathematics achievements (Mundy & Gilmore, 2009).

There are several aspects to consider when training the mapping between digits and quantities through a numerical estimation task. Arabic digits are a precise representation of quantities; so, an improvement in the comprehension of their meaning and the knowledge of the cardinality they represent was the purpose of this training. What the training may do is to help children to think of a specific magnitude when they think about numbers – to give them the clear idea that the more precisely one gets the mapping between a digit and a quantity, the better off she is when doing mathematics.

The training can improve children in the Operations test (3 ☐ 5 = 15) because, just as ANS training, it also accomplishes the mission of getting children used to think about quantities as the meanings of digits. Besides, individual children's acquisition of cardinal principle is related to an improvement in ANS acuity (Shusterman et al., 2016).

When the problem to solve consists in exact calculation (additions and subtractions tests) where an exact answer is required, is when the positive effects of training with digits and quantities could be evidenced. In the Additions and Subtractions test (Figures 1a and 1b; Studies 1 and 2), results indicate that the Numerical Estimation training was more effective than the Quantity Discrimination training for generating significant improvements additions and subtractions performance after training (Figure 9, Study 2). At 7-8 years of age, children already know how to add, subtract and multiply (tables 1 to 5 and 10). Considering that in neither training any arithmetic was taught, why a training aimed at giving children a better grasp of the quantities associated to digits, transfers into an improvement in their ability to calculate exact additions and subtractions, in a way in which simply focusing on ANS does not?

Taken as linguistic objects, Arabic digits are arbitrary words, as arbitrary as any other word in a natural language. Thus, the relation between Arabic digits and quantities has to be learned in order to master the mathematical language just as the relation between words and objects has to be learned in order to master a natural language. However, in the case of digits and quantities above the range of subitizing, could be necessary some practice to improve the calibration, although it can be achieved (Izard & Dehaene, 2008). Probably, it is not essential a perfect and exact association digit-quantity, but the more accurate the mapping the better (Mundy & Gilmore, 2009).

We may imagine that this training could rely on cognitive mechanism exploiting specific brain networks, partly endogenous and partly generated during the process of learning. The cognitive mechanism could be recruiting (in addition to the intraparietal region), a distinct left-hemisphere circuit associated with mathematic language for storing and retrieving math symbols and arithmetic facts, build on education-specific strategies (Dehaene, 2011). Indeed, crucially, the functional specialization of this left parietal circuit develops with schooling experience across a wide age span (Ansari & Dhital, 2006; Pinel & Dehaene, 2010; Rivera et al., 2005), so again practice could lead to improvement. Perhaps our training regime acts to the establishment of these circuits, and this in turn could lead to generate the improvements we found.

The basic arithmetic operations can be learned by mechanically executing calculation algorithms: just follow the instructions. These instructions, such as: “carry one…”, or “plus the one carried over…”… may be helpful to arrive at the correct result of one operation but lead children away from the real meaning of what they are doing. Children also learn timetables (and sometimes even basic additions and subtractions) by heart. Per se, these “school strategies” are not a problem. The problem is when they are used without a full comprehension of what they mean. This could easily happen for some children. For them, repeating the basic arithmetic operations over and over will not be of any help to solve the problem of understanding the meaning of what they are doing. This memory-based learning process does not improve their mathematical competence. It is not surprising then that lack of motivation (Simzar et al., 2016), predisposition (Cerda et al., 2015), and math anxiety (Pletzer et al., 2015; Z. Wang et al., 2014) come into play. Meanwhile, other students may somehow understand the mathematical language and the meaning of the strategies; for them, the practice of arithmetic takes on a different value, further helping to improve their mathematical competence. And so, a growing gap may be created between children who are "good at math" and those who are not. Changing this self-concept at later stages, will not be so easy.

Improving arithmetic skills, by reinforcing (or even creating) the relation between Arabic digits and the quantities they represent, can help to see the meaning of the calculation strategies that are learned at school. It can contribute to see mathematical calculations for what they are, to better “see” the objects that they manipulate.

We submit that the educational systems overestimate 7-8-year-olds' comprehension of this basic aspect of mathematical language. An appropriate training of this relation may complement standard school teaching, potentially generating long-lasting benefits in children's mathematical abilities and self-confidence.

5.3 ANS, Digit-Quantity mapping and School performance: concluding remarks.

In Studies 1 and 2, we have centered our attention to 7-8-year-old children. We trained their Approximate Number System by a quantity discrimination task. We have shown some of its advantages and its limits. The improvement it provides was not general and was not as solid as one would like to have it. Because mathematics has its own language, to improve math performance we need to "speak" in the language of numbers. This is a precise language, where approximation has no place. The Arabic digits reflect this precision. We thought that, for children to improve in mathematics, it was necessary to link their ANS with that language in the most precise and unequivocal way possible. This was the underlying motivation of our study 2. With a numerical estimation training (Digits computer program), children had to map Arabic digits to quantities. Thus, without using any indirect task with digits such as naming, counting or ordering, we obtained considerable benefits in arithmetic performance for all children of the trained age.

Once these benefits established, many questions come to the mind. Our study 3 was an attempt at finding at least some of them. They fall under two main broad topics: how numerical cognition develops and how numerical cognition relates with mathematics as practiced at school.

The development of numerical cognition

In study 3, we verified that ANS accuracy increases with age, at least from 8 to 16 year (Halberda & Feigenson, 2008).

Taking this result together with Study 1 and 2, we can extend this range from 7 to 16 years. As we showed, also the accuracy in the digits-quantities mapping improves with age, at least from 7 to 16 years.

We also found a remarkable speed-accuracy tradeoff both numerical cognition tasks, across all ages (8 to 16). When the students took more time to respond, their percentage of correctness increased. These findings suggest some interesting conclusions about the development of numerical cognition we should consider. In the field of research and proposing it as a topic to be improved in future investigations: a) the interpretation that small response times are better could be wrong; b) It could be an erroneous strategy to give participants instructions that both, speed and accuracy were important. In the field of education and specifically in mathematics, this may be the basis for pedagogical guidelines, suggesting that being quicker at responding should not be encouraged. Quite the opposite, teacher should insist on the importance of giving a correct answer regardless of how long it may take. Grasping the meaning should be prized over acquiring a skill.

Numerical cognition and mathematics at school

The most important relation for its implications for the educational word is the relation between numerical cognition and mathematics at school. Our measure of school performance is very reliable and conservative. We could check school performance in two successive quarters of the school year, and our tests was run between them.

The ability to map quantities into digits correlated positively with school math grades for a continuous and wide range of ages, from grade 3rd to grade 7th (8-to-13 years). A good knowledge of cardinality, that is of the quantities digit represent, is something relatively easy to achieve, but it may cause significant improvements in children's mathematical competence. Adding the results of study 2, we can conclude that the range of ages in which could be important to train the digit-quantity mapping extend from

7 to 13 years. These ages are crucial to establish the foundations for a better learning in advanced mathematics, as well as to give children confidence about their own possibilities.

Having a more precise ANS correlated with school grades, but only at some ages. The impact of training this ability to improve math competence in students is less marked than training the digit-quantity relations. It can still be useful in an indirect way, because we know that this ability correlates with the knowledge of the digit-quantity mapping. Thus, it could still be useful in order to help young children and low-profile students, especially because training it does not require the added difficulty of knowing any symbolic mathematical language.

5.4 Future directions

The research developed in this thesis has shed light on some questions concerning the relation between some basic mathematical abilities and mathematical performance. At the same time, other questions and hypothesis have come to the fore. With the firm commitment to make our research useful to the educational system, we propose some future directions of development.

One limitation of our studies is that our participants mainly came from middle-to-high socioeconomic environments. However, we found that some of the training regimes we tested were particularly useful to low-performing children. We believe that the training systems explored in this thesis have the potential of being most effective in less fortunate background, where stronger support and alternative corollary activities are most needed. We would like to confirm this hypothesis.

A second line of research that we would like to explore concerns the relation between mathematical abilities and dyslexia. Dyslexic individuals also have serious deficits in simple arithmetic (Callens, Tops, & Brysbaert, 2012; De Smedt & Boets, 2010; Evans, Flowers, Napoliello, Olulade, & Eden, 2014). Perhaps a training particularly centered on strengthening the relation between symbols and quantities, such as the Digits computer game, may have particularly beneficial effects among those who have the most difficulties in relating written symbols and their meanings in the development of mathematical language.

Finally, one very important fact that almost any research in developmental cognition has stressed is the centrality of early intervention. Biases and failures in mathematical performance appear very early, for example math anxiety (Maloney & Beilock, 2012). Currently, we have no data that address the question of how a specific training of the digit-quantity relation, appropriately adapted to earlier ages (3 to 6 year olds) could influence the understanding of activities with digits such as counting or ordering, and of the very basic mathematical facts. Even here, our research opens at the same time some novel perspectives and shows how much still needs to be done in order to understand how cognitively inspired interventions could affect learning and development, for the benefit of children and society.

References

Abreu-Mendoza, R. A., Soto-Alba, E. E., & Arias-Trejo, N. (2013). Area vs. density: influence of visual variables and cardinality knowledge in early number comparison. *Frontiers in Psychology*, *4*, 805. http://doi.org/10.3389/fpsyg.2013.00805

Agrillo, C., Piffer, L., & Bisazza, A. (2011). Number versus continuous quantity in numerosity judgments by fish. *Cognition*, *119*(2), 281–287. http://doi.org/10.1016/j.cognition.2010.10.022

Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences*, 201603205. http://doi.org/10.1073/pnas.1603205113

Ansari, D., & Dhital, B. (2006). Age-related Changes in the Activation of the Intraparietal Sulcus during Nonsymbolic Magnitude Processing: An Event-related Functional Magnetic Resonance Imaging Study. *Journal of Cognitive Neuroscience*, *18*(11), 1820–1828. http://doi.org/10.1162/jocn.2006.18.11.1820

Ansari, D., Donlan, C., Thomas, M. S. C., Ewing, S. A., Peen, T., & Karmiloff-Smith, A. (2003). What makes counting count? Verbal and visuo-spatial contributions to typical and atypical number development. *Journal of Experimental Child Psychology*, *85*(1), 50–62. http://doi.org/10.1016/S0022-0965(03)00026-2

Ansari, D., Garcia, N., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *NeuroReport*, *16*(16), 1769–1773. http://doi.org/10.1097/01.wnr.0000183905.23396.f1

Antell, S. E., & Keating, D. P. (1983). Perception of Numerical Invariance in Neonates. *Child Development*, *54*(3), 695. http://doi.org/10.2307/1130057

Arnaud, V., Hubbard, E. M., Dehaene, S., & Sackur, J. (2010). Number line compression and the illusory perception of random numbers. *Experimental Psychology*, *57*(6), 446–454. http://doi.org/10.1027/1618-3169/a000055

Arrighi, R., Togoli, I., & Burr, D. C. (2014). A generalized sense of number. *Proceedings. Biological Sciences / The Royal Society*, *281*(1797), 20141791-. http://doi.org/10.1098/rspb.2014.1791

Aubrey, C., Godfrey, R., & Dahl, S. (2006). Early mathematics development and later achievement: Further evidence. *Mathematics Education Research Journal*, *18*(1), 27–46. http://doi.org/10.1007/BF03217428

Banks, W. P., & Hill, D. K. (1974). The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology*, *102*(2), 353–376. http://doi.org/10.1037/h0035850

Benoit, L., Lehalle, H., & Jouen, F. (2004). Do young children acquire number words through subitizing or counting? *Cognitive Development*, *19*(3), 291–307. http://doi.org/10.1016/j.cogdev.2004.03.005

Beran, M. J., Evans, T. A., & Harris, E. H. (2008). Perception of food amounts by chimpanzees based on the number, size, contour length and visibility of items. *Animal Behaviour*, *75*(5), 1793–1802. http://doi.org/10.1016/j.anbehav.2007.10.035

Berch, D. B., Foley, E. J., Hill, R. J., & Ryan, P. M. (1999). Extracting parity and magnitude from Arabic numerals: developmental changes in number processing and mental representation. *Journal of Experimental Child Psychology*, *74*(4), 286–308. http://doi.org/10.1006/jecp.1999.2518

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., & Zorzi, M. (2010). Numerical estimation in preschoolers. *Developmental Psychology*, *46*(2), 545–51. http://doi.org/10.1037/a0017887

Bisazza, A., Piffer, L., Serena, G., & Agrillo, C. (2010). Ontogeny of numerical abilities in fish. *PLoS ONE*, *5*(11). http://doi.org/10.1371/journal.pone.0015516

Bonny, J. W., & Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: evidence from the preschool years. *Journal of Experimental Child Psychology*, *114*(3), 375–88. http://doi.org/10.1016/j.jecp.2012.09.015

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*(1), 189–201. http://doi.org/10.1037/0012-1649.41.6.189

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, *79*(4), 1016–1031. http://doi.org/10.1111/j.1467-8624.2008.01173.x

Brannon, E. M. (2006). The representation of numerical magnitude. *Current Opinion in Neurobiology*, *16*(2), 222–229.

http://doi.org/10.1016/j.conb.2006.03.002
Bugden, S., & Ansari, D. (2015). Probing the nature of deficits in the “Approximate Number System” in children with persistent Developmental Dyscalculia. *Developmental Science*. http://doi.org/10.1111/desc.12324
Bugden, S., DeWind, N. K., & Brannon, E. M. (2016). Using cognitive training studies to unravel the mechanisms by which the approximate number system supports symbolic math ability. *Current Opinion in Behavioral Sciences*, *10*, 73–80. http://doi.org/10.1016/j.cobeha.2016.05.002
Bulf, H., de Hevia, M. D., Gariboldi, V., & Macchi Cassia, V. (2017). Infants learn better from left to right: a directional bias in infants’ sequence learning. *Scientific Reports*, *7*(1), 2437. http://doi.org/10.1038/s41598-017-02466-w
Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences*, *14*(12), 534–41. http://doi.org/10.1016/j.tics.2010.09.007
Callens, M., Tops, W., & Brysbaert, M. (2012). Cognitive profile of students Who enter higher education with an indication of Dyslexia. *PLoS ONE*, *7*(6), e38081. http://doi.org/10.1371/journal.pone.0038081
Cantlon, J. F., Brannon, E. M., Carter, E. J., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biology*, *4*(5), e125. http://doi.org/10.1371/journal.pbio.0040125
Cantlon, J. F., Libertus, M. E., Pinel, P., Dehaene, S., Brannon, E. M., & Pelphrey, K. A. (2009). The Neural Development of an Abstract Concept of Number. *Journal of Cognitive Neuroscience*, *21*(11), 2217–2229. http://doi.org/10.1162/jocn.2008.21159
Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Sciences*, *13*(2), 83–91. http://doi.org/10.1016/j.tics.2008.11.007
Carey, S. (2004). *Bootstrapping & the origin of concepts*. *Daedalus* (Vol. 133). http://doi.org/10.1162/001152604772746701
Castronovo, J., & Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PloS One*, *7*(4), e33832. http://doi.org/10.1371/journal.pone.0033832
Cerda, G., Pérez, C., Navarro, J. I., Aguilar, M., Casas, J. a, & Aragón, E. (2015). Explanatory model of emotional-cognitive variables in school mathematics performance: a longitudinal

study in primary school. *Frontiers in Psychology*, *6*(September), 1363. http://doi.org/10.3389/fpsyg.2015.01363

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica, 148,* 163–172. http://doi.org/10.1016/j.actpsy.2014.01.016

Clayton, S., & Gilmore, C. (2014). Inhibition in dot comparison tasks. *Zdm*, 1–12. http://doi.org/10.1007/s11858-014-0655-2

Clayton, S., Gilmore, C., & Inglis, M. (2015). Dot comparison stimuli are not all alike: the effect of different visual controls on ANS measurement. *Acta Psychologica*, *161*, 177–84. http://doi.org/10.1016/j.actpsy.2015.09.007

Coch, D., & Ansari, D. (2009). Thinking about mechanisms is crucial to connecting neuroscience and education. *Cortex*, *45*(4), 546–547. http://doi.org/10.1016/j.cortex.2008.06.001

Cohen, D. J., & Sarnecka, B. W. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental Psychology*, *50*(6), 1640–1652. http://doi.org/10.1037/a0035901

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin {&} Review*, *8*(4), 698–707. http://doi.org/10.3758/BF03196206

Coubart, A., Izard, V., Spelke, E. S., Marie, J., & Streri, A. (2014). Dissociation between small and large numerosities in newborn infants. *Developmental Science*, *17*(1), 11–22. http://doi.org/10.1111/desc.12108

Dadda, M., Piffer, L., Agrillo, C., & Bisazza, A. (2009). Spontaneous number representation in mosquitofish. *Cognition*, *112*(2), 343–348. http://doi.org/10.1016/j.cognition.2009.05.009

de Hevia, M. D., Izard, V., Coubart, A., Spelke, E. S., & Streri, A. (2014). Representations of space, time, and number in neonates. *Proceedings of the National Academy of Sciences*, *111*(13), 4809–4813. http://doi.org/10.1073/pnas.1323628111

de Hevia, M. D., Vallar, G., & Girelli, L. (2008). Visualizing numbers in the mind's eye: The role of visuo-spatial processes in numerical abilities. *Neuroscience and Biobehavioral Reviews*, *32*(8), 1361–1372. http://doi.org/10.1016/j.neubiorev.2008.05.015

De Smedt, B., & Boets, B. (2010). Phonological processing and arithmetic fact retrieval: Evidence from developmental dyslexia. *Neuropsychologia*, *48*(14), 3973–3981. http://doi.org/10.1016/j.neuropsychologia.2010.10.018

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1–2), 1–42. http://doi.org/10.1016/0010-0277(92)90049-N

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 & Language*, *16*(1), 16–36. http://doi.org/10.1111/1468-0017.00154

Dehaene, S. (2007). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In P. Haggard & Y. Rossetti (Eds.), *Attention & Performance XXII. Sensori-motor foundations of higher cognition.* (pp. 527–574). Cambridge, Mass.: Harvard Universitiy Press. http://doi.org/10.1093/acprof:oso/9780199231447.003.0024

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(2), 314–326. http://doi.org/10.1037/0278-7393.21.2.314

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396. http://doi.org/10.1037/0096-3445.122.3.371

Dehaene, S., & Cohen, L. (1991). Two mental calculation systems: a case study of severe acalculia with preseved approximation. *Neuropsychologia*, *29*(11), 1045–1074. http://doi.org/10.1016/0028-3932(91)90076-K

Dehaene, S., & Cohen, L. (1997). Cerebral Pathways for Calculation: Double Dissociation between Rote Verbal and Quantitative Knowledge of Arithmetic. *Cortex*, *33*(2), 219–250. http://doi.org/10.1016/S0010-9452(08)70002-9

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neurosciences*, *21*(8), 355–361. http://doi.org/10.1016/S0166-2236(98)01263-6

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental*

Psychology: Human Perception and Performance, *16*(3), 626–641. http://doi.org/10.1037/0096-1523.16.3.626
Dehaene, S., Izard, V., Spelke, E. S., & Pica, P. (2008). Log or Linear? Distinct Intuitions of the Number Scale in Western and Amazonian Indigene Cultures. *Science*, *320*(5880), 1217–1220. http://doi.org/10.1126/science.1156540
Dehaene, S., & Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, *43*(1), 1–29. http://doi.org/10.1016/0010-0277(92)90030-L
Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition*, *79*(1–2), 1–37. http://doi.org/10.1016/S0010-0277(00)00123-2
Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., … Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), 597–600. http://doi.org/10.1038/26967
Demaray, M. K., & Elliot, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, *13*(1), 8–24. http://doi.org/10.1037/h0088969
den Heyer, K., & Briand, K. (1986). Priming single digit numbers: Automatic spreading activation dissipates as a function of semantic distance. *The American Journal of Psychology*, *99*(3), 315–340. http://doi.org/10.2307/1422488
Desimone, L., & Long, D. (2010). Teacher effects and the achievement gap: Do teacher and teaching quality influence the achievement gap between black and white and high-and low-SES. *Teachers College Record*. Retrieved from http://www.gse.upenn.edu/pdf/desimone/ECLS TCR.pdf
Desoete, A., & Praet, M. (2013). Inclusive mathematics education: The value of a computerized look-ahead approach in kindergarten. A randomized controlled study. *Special Issue: Psychological and Educational Aspects of Inclusion.*, *Spec Issue*, 103–119. Retrieved from http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc10&NEWS=N&AN=2014-23573-006
Dewind, N. K., & Brannon, E. M. (2012). Malleability of the approximate number system: effects of feedback and training. *Frontiers in Human Neuroscience*, *6*, 68. http://doi.org/10.3389/fnhum.2012.00068

Dyson, N., Jordan, N. C., Beliakoff, A., & Hassinger-Das, B. (2015). A Kindergarten Number-Sense Intervention With Contrasting Practice Conditions for Low-Achieving Children. *Journal for Research in Mathematics Education*, *46*(3), 331. http://doi.org/10.5951/jresematheduc.46.3.0331

Elena, M., Petrazzini, M., Agrillo, C., Izard, V., & Bisazza, A. (2016). Do humans (Homo sapiens) and fish (Pterophyllum scalare) make similar numerosity judgments ? *Journal of Comparative Psychology*, *130*(4), 380–390. http://doi.org/10.1037/com0000045

Emerson, R. W., & Cantlon, J. F. (2015). Continuity and change in children's longitudinal neural responses to numbers. *Developmental Science*, *18*(2), 314–26. http://doi.org/10.1111/desc.12215

Evans, T. M., Flowers, D. L., Napoliello, E. M., Olulade, O. A., & Eden, G. F. (2014). The functional anatomy of single-digit arithmetic in children with developmental dyslexia. *NeuroImage*, *101*, 644–652. http://doi.org/10.1016/j.neuroimage.2014.07.028

Feigenson, L., Dehaene, S., & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. http://doi.org/10.1016/j.tics.2004.05.002

Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links Between the Intuitive Sense of Number and Formal Mathematics Ability. *Child Development Perspectives*, *7*(2), 74–79. http://doi.org/10.1111/cdep.12019

Ferres-Forga, N., Bonatti, L. L., & Halberda, J. (2017). One-Month of Approximate Number System Training Improves Symbolic Mathematics Performance, But Only In Low-Performing Children. *Submitted*.

Ferres-Forga, N., Halberda, J., & Bonatti, L. L. (2017). Improving Mathematics Performance in 7-year-old children: training the Mapping from Quantity to Digit. *Submitted*.

Fias, W., Reynvoet, B., & Brysbaert, M. (2001). Are Arabic numerals processed as pictures in a Stroop interference task? *Psychological Research*, *65*(4), 242–249. http://doi.org/10.1007/s004260100064

Fuhs, M. W., & McNeil, N. M. (2015). The Role of Non-Numerical Stimulus Features in Approximate Number System Training in Preschoolers from Low-Income Homes Mary. *Statewide Agricultural Land Use Baseline 2015*, *1*(June), 1–58.

http://doi.org/10.1017/CBO9781107415324.004
Geary, D. C. (2013). Early Foundations for Mathematics Learning and Their Relations to Learning Disabilities. *Curr Dir Psychol Sci.*, *22*(1), 43–51. http://doi.org/10.1177/0963721412469398.Early
Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' Functional Numeracy Is Predicted by Their School Entry Number System Knowledge. *PLoS ONE*, *8*(1). http://doi.org/10.1371/journal.pone.0054651
Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, Mass.: Harvard Universitiy Press.
Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequence is spatially organized. *Cognition*, *87*, B87–B95. http://doi.org/10.1016/S0010-9452(08)70938-9
Gevers, W., Reynvoet, B., & Fias, W. (2004). The Mental Representation of Ordinal Sequences is Spatially Organised: Evidence from Days of the Week. *Cortex*, *40*(1), 171–172. http://doi.org/10.1016/S0010-9452(08)70938-9
Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition*, *115*(3), 394–406. http://doi.org/10.1016/j.cognition.2010.02.002
Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *Journal of Experimental Child Psychology*, *76*(2), 104–122. http://doi.org/10.1006/jecp.2000.2564
Goffin, C., & Ansari, D. (2016). Beyond magnitude: Judging ordinality of symbolic number is unrelated to magnitude comparison and independently relates to individual differences in arithmetic. *Cognition*, *150*, 68–76. http://doi.org/10.1016/j.cognition.2016.01.018
González, M., Kittredge, A., Sánchez, I., Fleischer, B., Spelke, E. S., & Maiche, A. (2015). A way to improve math skills through stimulating ANS. Retrieved December 7, 2015, from http://www.cognicionnumerica.psico.edu.uy/wp-content/uploads/2015/06/articulo_neuroeducacao.pdf
Gordon, P. (2004). Numerical Cognition Without Words: Evidence from Amazonia. *Science*, *306*(5695), 496–499. http://doi.org/10.1126/science.1094492
Gray, S. A., & Reeve, R. A. (2014). Preschoolers' dot enumeration

abilities are markers of their arithmetic competence. *PLoS ONE*, *9*(4). http://doi.org/10.1371/journal.pone.0094428

Guillaume, M., Gevers, W., & Content, A. (2016). Assessing the Approximate Number System: no relation between numerical comparison and estimation tasks. *Psychological Research*, *80*(2), 248–58. http://doi.org/10.1007/s00426-015-0657-x

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), 1457–65. http://doi.org/10.1037/a0012682

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(28), 11116–20. http://doi.org/10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–8. http://doi.org/10.1038/nature07246

Haworth, C. M. A., Kovas, Y., Petrill, S. A., & Plomin, R. (2007). Developmental origins of low mathematics performance and normal variation in twins from 7 to 9 years. *Twin Research and Human Genetics : The Official Journal of the International Society for Twin Studies*, *10*(1), 106–117. http://doi.org/10.1375/twin.10.1.106

He, L., Zhou, K., Zhou, T., He, S., & Chen, L. (2015). Topology-defined units in numerosity perception. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41), E5647-55. http://doi.org/10.1073/pnas.1512408112

He, Y., Zhou, X., Shi, D., Song, H., Zhang, H., & Shi, J. (2016). New Evidence on Causal Relationship between Approximate Number System (ANS) Acuity and Arithmetic Ability in Elementary-School Students: A Longitudinal Cross-Lagged Analysis. *Frontiers in Psychology*, *7*(July), 1–8. http://doi.org/10.3389/fpsyg.2016.01052

Henik, A., & Tzelgov, J. (1982). Is three greater than five: The relation between physical and semantic size in comparison tasks. *Memory & Cognition*, *10*(4), 389–395. http://doi.org/10.3758/BF03202431

Hermelin, B., & O’Connor, N. (1986). Spatial representations in mathematically and in artistically gifted children. *British*

Journal of Educational Psychology, *56*(2), 150–157. http://doi.org/10.1111/j.2044-8279.1986.tb02656.x

Herrera, A., & Macizo, P. (2011). Naming digits in a semantic blocking paradigm. *Quarterly Journal of Experimental Psychology*, *64*(2), 328–338. http://doi.org/10.1080/17470218.2010.508532

Herrera, A., & Macizo, P. (2012). Semantic processing in the production of numerals across notations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 40–51. http://doi.org/10.1037/a0024884

Hinrichs, J. V, Yurko, D. S., & Hu, J. (1981). Two-digit number comparison: Use of place information. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(4), 890–901. http://doi.org/10.1037/0096-1523.7.4.890

Hollingsworth, W. H., Simmons, J. P., Coates, T. R., & Cross, H. A. (1991). Perceived numerosity as a function of array number, speed of array development, and density of array Items. *Bulletin of the Psychonomic Society*, *29*(5), 448–450. http://doi.org/10.3758/BF03333967

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews. Neuroscience*, *6*(6), 435—448. http://doi.org/10.1038/nrn1684

Hyde, D. C. (2011). Two systems of non-symbolic numerical cognition. *Frontiers in Human Neuroscience*, *5*, 150. http://doi.org/10.3389/fnhum.2011.00150

Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, *131*(1), 92–107. http://doi.org/10.1016/j.cognition.2013.12.007

Hyde, D. C., & Spelke, E. S. (2009). All Numbers Are Not Equal: An Electrophysiological Investigation of Small and Large Number Representations. *Journal of Cognitive Neuroscience*, *21*(6), 1039–1053. http://doi.org/10.1162/jocn.2009.21090

Hyde, D. C., & Spelke, E. S. (2011). Neural signatures of number processing in human infants: evidence for two core systems underlying numerical cognition. *Developmental Science*, *14*(2), 360–71. http://doi.org/10.1111/j.1467-7687.2010.00987.x

Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National*

Academy of Sciences of the United States of America, *106*(22), 8801–7. http://doi.org/10.1073/pnas.0901265106

Izard, V., Coralie, S., Elizabeth, S., Spelke, A., & Streri. (2009). Newborn Infants Perceive Abstract Numbers. *Proceedings of the National Academy of Sciences*, *106*(25), 10382–10385. http://doi.org/10.1073/pnas.0812142106

Izard, V., Dehaene-Lambertz, G., & Dehaene, S. (2008). Distinct cerebral pathways for object identity and number in human infants. *PLoS Biology*, *6*(2), 0275–0285. http://doi.org/10.1371/journal.pbio.0060011

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–47. http://doi.org/10.1016/j.cognition.2007.06.004

Jansen, B. R. J., Louwerse, J., Straatemeier, M., Van der Ven, S. H. G., Klinkenberg, S., & Van der Maas, H. L. J. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences*, *24*, 190–197. http://doi.org/10.1016/j.lindif.2012.12.014

Jones, S. M., & Brannon, E. M. (2012). Prosimian primates show ratio dependence in spontaneous quantity discriminations. *Frontiers in Psychology*, *3*, 550. http://doi.org/10.3389/fpsyg.2012.00550

Jones, S. M., Pearson, J., DeWind, N. K., Paulsen, D., Tenekedjieva, A.-M., & Brannon, E. M. (2014). Lemurs and macaques show similar numerical sensitivity. *Animal Cognition*, *17*(3), 503–15. http://doi.org/10.1007/s10071-013-0682-3

Jordan, N. C., Kaplan, D., Ramineni, C., & Locuniak, M. N. (2009). Early math matters: kindergarten number competence and later mathematics outcomes. *Developmental Psychology*, *45*(3), 850–67. http://doi.org/10.1037/a0014939

Jou, J., & Aldridge, J. W. (1999). Memory representation of alphabetic position and interval information. *Journal of Experimental Psychology Learning Memory and Cognition*, *25*(3), 680–701. http://doi.org/10.1037//0278-7393.25.3.680

Kemp, S. (1984). Estimating the Sizes of Sports Crowds. *Perceptual and Motor Skills*, *59*(3), 723–729. http://doi.org/10.2466/pms.1984.59.3.723

Knoll, L. J., Fuhrmann, D., Sakhardande, A. L., Stamp, F., Speekenbrink, M., & Blakemore, S.-J. (2016). A Window of

Opportunity for Cognitive Training in Adolescence. *Psychological Science*, 956797616671327. http://doi.org/10.1177/0956797616671327

Knops, A., Dehaene, S., Berteletti, I., & Zorzi, M. (2014). Can approximate mental calculation account for operational momentum in addition and subtraction? *Quarterly Journal of Experimental Psychology (2006)*, *67*(8), 1541–56. http://doi.org/10.1080/17470218.2014.890234

Kroesbergen, E., Luit, J. Van, & Maas, C. (2004). Effectiveness of explicit and constructivist mathematics instruction for low-achieving students in the Netherlands. *The Elementary School*. Retrieved from http://www.jstor.org/stable/3202951

Kroll, J. F., & Stewart, E. (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections Between Bilingual Memory Representations. *Journal of Memory and Language*, *33*(2), 149–174. http://doi.org/10.1006/jmla.1994.1008

Krueger, L. E. (1982). Single judgments of numerosity. *Perception {&} Psychophysics*, *31*(2), 175–182. http://doi.org/10.3758/BF03206218

Krueger, L. E. (1989). Reconciling Fechner and Stevens: Toward a unified psychophysical law. *Behavioral and Brain Sciences*, *12*(2), 251–267. http://doi.org/10.1017/S0140525X0004855X

Laski, E. V., & Siegler, R. S. (2007). Is 27 a big number? correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development*, *78*(6), 1723–1743. http://doi.org/10.1111/j.1467-8624.2007.01087.x

Le Corre, M. (2014). Children acquire the later-greater principle after the cardinal principle. *British Journal of Developmental Psychology*, *32*(2), 163–177. http://doi.org/10.1111/bjdp.12029

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395–438. http://doi.org/10.1016/j.cognition.2006.10.005

Leibovich, T., & Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Frontiers in Psychology*, *4*, 375. http://doi.org/10.3389/fpsyg.2013.00375

Lemer, C., Dehaene, S., Spelke, E. S., & Cohen, L. (2003). Approximate quantities and exact number words: dissociable systems. *Neuropsychologia*, *41*(14), 1942–58. Retrieved from

http://www.ncbi.nlm.nih.gov/pubmed/14572527
Libertus, M. E., & Brannon, E. M. (2009). Behavioral and Neural Basis of Number Sense in Infancy. *Current Directions in Psychological Science*, *18*(6), 346–351. http://doi.org/10.1111/j.1467-8721.2009.01665.x
Libertus, M. E., & Brannon, E. M. (2010). Stable individual differences in number discrimination in infancy. *Developmental Science*, *13*(6), 900–6. http://doi.org/10.1111/j.1467-7687.2009.00948.x
Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science*, *14*, 1292–300. http://doi.org/10.1111/j.1467-7687.2011.01080.x
Libertus, M. E., Feigenson, L., & Halberda, J. (2013a). Is Approximate Number Precision a Stable Predictor of Math Ability? *Learning and Individual Differences*, *25*, 126–133. http://doi.org/10.1016/j.lindif.2013.02.001
Libertus, M. E., Feigenson, L., & Halberda, J. (2013b). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *Journal of Experimental Child Psychology*, *116*(4), 829–38. http://doi.org/10.1016/j.jecp.2013.08.003
Libertus, M. E., Feigenson, L., Halberda, J., & Landau, B. (2014). Understanding the mapping between numerical approximation and number words: evidence from Williams syndrome and typical development. *Developmental Science*, *17*(6), 905–919. http://doi.org/10.1111/desc.12154
Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica*, *141*(3), 373–9. http://doi.org/10.1016/j.actpsy.2012.09.009
Lindskog, M., & Winman, A. (2016). No evidence of learning in non-symbolic numerical tasks - A comment on Park and Brannon (2014). *Cognition*, *150*, 243–247. http://doi.org/10.1016/j.cognition.2016.01.005
Linsen, S., Verschaffel, L., Reynvoet, B., & De Smedt, B. (2014). The association between children’s numerical magnitude processing and mental multi-digit subtraction. *Acta Psychologica*, *145*(1), 75–83. http://doi.org/10.1016/j.actpsy.2013.10.008
Lussier, C. A., & Cantlon, J. F. (2017). Developmental bias for

number words in the intraparietal sulcus. *Developmental Science*, *20*(3), e12385. http://doi.org/10.1111/desc.12385

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*, *121*(2), 256–61. http://doi.org/10.1016/j.cognition.2011.07.009

Lyons, I. M., & Beilock, S. L. (2013). Ordinality and the nature of symbolic numbers. *The Journal of Neuroscience*, *33*(43), 17052–61. http://doi.org/10.1523/JNEUROSCI.1775-13.2013

Maloney, E. A., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Sciences*, *16*(8), 404–406. http://doi.org/10.1016/j.tics.2012.06.008

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011a). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development*, *82*(4), 1224–1237. http://doi.org/10.1111/j.1467-8624.2011.01608.x

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011b). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, *6*(9), 1–9. http://doi.org/10.1371/journal.pone.0023749

McCrink, K., Spelke, E. S., Dehaene, S., & Pica, P. (2013). Non-symbolic halving in an Amazonian indigene group. *Developmental Science*, *16*(3), 451–62. http://doi.org/10.1111/desc.12037

Miletto Petrazzini, M. E., Agrillo, C., Izard, V., & Bisazza, A. (2015). Relative versus absolute numerical representation in fish: Can guppies represent "fourness"? *Animal Cognition*, *18*(5), 1007–17. http://doi.org/10.1007/s10071-015-0868-y

Moll, K., Snowling, M. J., Göbel, S. M., & Hulme, C. (2015). Early language and executive skills predict variations in number and arithmetic skills in children at family-risk of dyslexia and typically developing controls. *Learning and Instruction*, *38*, 53–62. http://doi.org/10.1016/j.learninstruc.2015.03.004

Moore, A. M., vanMarle, K., & Geary, D. C. (2016). Kindergartners' fluent processing of symbolic numerical magnitude is predicted by their cardinal knowledge and implicit understanding of arithmetic 2years earlier. *Journal of Experimental Child Psychology*, *150*, 31–47. http://doi.org/10.1016/j.jecp.2016.05.003

Morgan, P. L., Farkas, G., & Maczuga, S. (2015). Which Instructional Practices Most Help First Grade Students with and without Mathematics Difficulties? *HHS Public Access*, *37*(2), 184–205. http://doi.org/10.3102/0162373714536608

Mou, Y., & VanMarle, K. (2014). Two core systems of numerical representation in infants. *Developmental Review*, *34*(1), 1–25. http://doi.org/10.1016/j.dr.2013.11.001

Moyer, R. S., & Landauer, T. K. (1967). Time required for Judgements of Numerical Inequality. *Nature*, *215*(5109), 1519–1520. http://doi.org/10.1038/2151519a0

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology*, *103*(4), 490–502. http://doi.org/10.1016/j.jecp.2009.02.003

Naccache, L., & Dehaene, S. (2001a). The Priming Method: Imaging Unconscious Repetition Priming Reveals an Abstract Representation of Number in the Parietal Lobes. *Cerebral Cortex*, *11*(10), 966–974. http://doi.org/10.1093/cercor/11.10.966

Naccache, L., & Dehaene, S. (2001b). Unconscious semantic priming extends to novel unseen stimuli. *Cognition*, *80*(3), 223–237. http://doi.org/10.1016/S0010-0277(00)00139-6

Nieder, A., & Dehaene, S. (2009). Representation of Number in the Brain. *Annual Review of Neuroscience*, *32*(1), 185–208. http://doi.org/10.1146/annurev.neuro.051508.135550

Niederhauser, D. ., & Stoddart, T. (2001). Teachers' instructional perspectives and use of educational software. *Teaching and Teacher Education*, *17*(1), 15–31. http://doi.org/10.1016/S0742-051X(00)00036-6

Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PloS One*, *8*(7), e67918. http://doi.org/10.1371/journal.pone.0067918

Núñez-Peña, M. I., & Suárez-Pellicioni, M. (2014). Less precise representation of numerical magnitude in high math-anxious individuals: An ERP study of the size and distance effects. *Biological Psychology*, *103*, 176–183. http://doi.org/10.1016/j.biopsycho.2014.09.004

Núñez-Peña, M. I., Suárez-Pellicioni, M., & Bono, R. (2013).

Effects of math anxiety on student success in higher education. *International Journal of Educational Research*, *58*, 36–43. http://doi.org/10.1016/j.ijer.2012.12.004

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction*, *23*(1), 125–135. http://doi.org/10.1016/j.learninstruc.2012.08.004

Odic, D., Hock, H., & Halberda, J. (2012). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology. General*, *142*(1), 255–65. http://doi.org/10.1037/a0030825

Odic, D., Im, H. Y., Eisinger, R., Ly, R., & Halberda, J. (2015). PsiMLE: A maximum-likelihood estimation approach to estimating psychophysical scaling and variability more reliably, efficiently, and flexibly. *Behavior Research Methods*. http://doi.org/10.3758/s13428-015-0600-5

Odic, D., Le Corre, M., & Halberda, J. (2015). Children's mappings between number words and the approximate number system. *Cognition*, *138*, 102–21. http://doi.org/10.1016/j.cognition.2015.01.008

Papay, J. P., Murnane, R. J., & Willett, J. B. (2014). High-School Exit Examinations and the Schooling Decisions of Teenagers: Evidence From Regression-Discontinuity Approaches. *Journal of Research on Educational Effectiveness*, *7*(1), 1–27. http://doi.org/10.1080/19345747.2013.819398

Park, J., Bermudez, V., Roberts, R. C., & Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *Journal of Experimental Child Psychology*, *152*, 278–293. http://doi.org/10.1016/j.jecp.2016.07.011

Park, J., & Brannon, E. M. (2014). Improving arithmetic performance with number sense training: an investigation of underlying mechanism. *Cognition*, *133*(1), 188–200. http://doi.org/10.1016/j.cognition.2014.06.011

Park, J., & Brannon, E. M. (2016). How to interpret cognitive training studies: A reply to Lindskog & Winman. *Cognition*, *150*, 247–251. http://doi.org/10.1016/j.cognition.2016.02.012

Park, J., & Starns, J. J. (2015). The approximate number system acuity redefined: A diffusion model approach. *Frontiers in Psychology*, *6*(DEC), 1–10.

http://doi.org/10.3389/fpsyg.2015.01955
Parsons, S., & Bynner, J. (2005). *Does numeracy matter more?* London: National Research and Development Centre for Adult Literacy and Numeracy. Retrieved from http://dera.ioe.ac.uk/22317/1/doc_2979.pdf
Patalano, A. L., Saltiel, J. R., Machlin, L., & Barth, H. (2015). The role of numeracy and approximate number system acuity in predicting value and probability distortion. *Psychonomic Bulletin & Review*, *22*(6), 1820–9. http://doi.org/10.3758/s13423-015-0849-9
Peters, E., Hart, P. S., Tusler, M., & Fraenkel, L. (2014). Numbers matter to informed patient choices: a randomized design across age and numeracy levels. *Medical Decision Making : An International Journal of the Society for Medical Decision Making*, *34*(4), 430–42. http://doi.org/10.1177/0272989X13511705
Piazza, M., Giacomini, E., Le Bihan, D., & Dehaene, S. (2003). Single-trial classification of parallel pre-attentive and serial attentive processes using functional magnetic resonance imaging. *Proceedings of the Royal Society B: Biological Sciences*, *270*(1521), 1237–1245. http://doi.org/10.1098/rspb.2003.2356
Piazza, M., & Izard, V. (2009). How humans count: numerosity and the parietal cortex. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *15*(3), 261–73. http://doi.org/10.1177/1073858409333073
Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555. http://doi.org/10.1016/j.neuron.2004.10.014
Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychological Science*, *24*(6), 1037–43. http://doi.org/10.1177/0956797612464057
Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A Magnitude Code Common to Numerosities and Number Symbols in Human Intraparietal Cortex. *Neuron*, *53*(2), 293–305. http://doi.org/10.1016/j.neuron.2006.11.022
Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science (New York, N.Y.)*, *306*(5695), 499–503.

http://doi.org/10.1126/science.1102085
Pinel, P., & Dehaene, S. (2010). Beyond Hemispheric Dominance: Brain Regions Underlying the Joint Lateralization of Language and Arithmetic to the Left Hemisphere. *Journal of Cognitive Neuroscience*, *22*(1), 48–66. http://doi.org/10.1162/jocn.2009.21184
Pinel, P., Dehaene, S., Rivière, D., & LeBihan, D. (2001). Modulation of Parietal Activation by Semantic Distance in a Number Comparison Task. *NeuroImage*, *14*(5), 1013–1026. http://doi.org/10.1006/nimg.2001.0913
Pinheiro-Chagas, P., Wood, G., Knops, A., Krinzinger, H., Lonnemann, J., Starling-Alves, I., … Haase, V. G. (2014). In how many ways is the approximate number system associated with exact calculation? *PloS One*, *9*(11), e111155. http://doi.org/10.1371/journal.pone.0111155
Pletzer, B., Kronbichler, M., Nuerk, H. C., & Kerschbaum, H. H. (2015). Mathematics anxiety reduces default mode network deactivation in response to numerical tasks. *Frontiers in Human Neuroscience*, *9*(April), 1–12. http://doi.org/10.3389/fnhum.2015.00202
Prado, J., Mutreja, R., Zhang, H., Mehta, R., Desroches, A. S., Minas, J. E., & Booth, J. R. (2011). Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. *Human Brain Mapping*, *32*(11), 1932–1947. http://doi.org/10.1002/hbm.21159
Praet, M., & Desoete, A. (2014). Enhancing young children's arithmetic skills through non-intensive, computerised kindergarten interventions: A randomised controlled study. *Teaching and Teacher Education*, *39*, 56–65. http://doi.org/10.1016/j.tate.2013.12.003
Purpura, D. J., & Logan, J. A. R. (2015). The nonlinear relations of the approximate number system and mathematical language to early mathematics development. *Developmental Psychology*, *51*(12), 1717–24. http://doi.org/10.1037/dev0000055
Ramani, G. B., & Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. *Journal of Applied Developmental Psychology*, *32*(3), 146–159. http://doi.org/10.1016/j.appdev.2011.02.005
Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does Subitizing Reflect Numerical Estimation? *Psychological Science*, *19*(6), 607–614.

http://doi.org/10.1111/j.1467-9280.2008.02130.x
Reynvoet, B., & Ratinckx, E. (2004). Hemispheric differences between left and right number representations: Effects of conscious and unconscious priming. *Neuropsychologia*, *42*(6), 713–726. http://doi.org/10.1016/j.neuropsychologia.2003.11.013
Ritchie, S. J., & Bates, T. C. (2013). Enduring Links From Childhood Mathematics and Reading Achievement to Adult Socioeconomic Status. *Psychological Science*, *24*(7), 1301–1308. http://doi.org/10.1177/0956797612466268
Rivera, S. M., Reiss, A. L., Eckert, M. A., & Menon, V. (2005). Developmental changes in mental arithmetic: Evidence for increased functional specialization in the left inferior parietal cortex. *Cerebral Cortex*, *15*(11), 1779–1790. http://doi.org/10.1093/cercor/bhi055
Roelofs, A. (2006). Functional architecture of naming dice, digits, and number words. *Language and Cognitive Processes*, *21*(1–3), 78–111. http://doi.org/10.1080/01690960400001846
Rouder, J. N., & Geary, D. C. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science*, *17*(4), 525–36. http://doi.org/10.1111/desc.12166
Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *Journal of Experimental Child Psychology*, *81*(1), 74–92. http://doi.org/10.1006/jecp.2001.2645
Rugani, R., & de Hevia, M.-D. (2017). Number-space associations without language: Evidence from preverbal human infants and non-human animal species. *Psychonomic Bulletin & Review*, *24*(2), 352–369. http://doi.org/10.3758/s13423-016-1126-2
Sarnecka, B. W., & Carey, S. (2008). How counting represents number: What children must learn and when they learn it. *Cognition*, *108*(3), 662–674. http://doi.org/10.1016/j.cognition.2008.05.007
Sarnecka, B. W., & Lee, M. D. (2009). Levels of number knowledge during early childhood. *Journal of Experimental Child Psychology*, *103*(3), 325–337. http://doi.org/10.1016/j.jecp.2009.02.007
Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics

achievement. *British Journal of Developmental Psychology*, *30*(2), 344–357. http://doi.org/10.1111/j.2044-835X.2011.02048.x

Sasanguie, D., Defever, E., Maertens, B., & Reynvoet, B. (2014). The approximate number system is not predictive for symbolic number processing in kindergarteners. *Quarterly Journal of Experimental Psychology (2006)*, *67*(2), 271–80. http://doi.org/10.1080/17470218.2013.803581

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology*, *7*(1), 82–138. http://doi.org/10.1016/0010-0285(75)90006-7

Shusterman, A., Slusser, E., Halberda, J., & Odic, D. (2016). Acquisition of the cardinal principle coincides with improvement in approximate number system acuity in preschoolers. *PLoS One*, *11*(4), e0153072. http://doi.org/10.1371/journal.pone.0153072

Siegler, R. S. (2003). Implications of cognitive science research for mathematics education. *A Research Companion to Principles and Standards for School Mathematics*, 219–233.

Siegler, R. S., & Robinson, M. (1982). The Development of Numerical Understandings. *Advances in Child Development and Behavior*, *16*(C), 241–312. http://doi.org/10.1016/S0065-2407(08)60072-5

Simzar, R. M., Domina, T., & Tran, C. (2016). Eighth Grade Algebra Course Placement and Student Motivation for Mathematics. *HHS Public Access*, *2*(1), 1–26. http://doi.org/10.1177/2332858415625227

Sokolowski, H. M., Fias, W., Mousa, A., & Ansari, D. (2017). Common and distinct brain regions in both parietal and frontal cortex support symbolic and nonsymbolic number processing in humans: A functional neuroimaging meta-analysis. *NeuroImage*, *146*, 376–394. http://doi.org/10.1016/j.neuroimage.2016.10.028

Spelke, E. S., & Tsivkin, S. (2001). Language and number: a bilingual training study. *Cognition*, *78*(1), 45–88. http://doi.org/10.1016/S0010-0277(00)00108-6

Starr, A. B., Libertus, M. E., & Brannon, E. M. (2013). Infants Show Ratio-dependent Number Discrimination Regardless of Set Size. *Infancy : The Official Journal of the International Society on Infant Studies*, *18*(6). http://doi.org/10.1111/infa.12008

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(45), 18116–20. http://doi.org/10.1073/pnas.1302751110

Stoet, G., Bailey, D. H., Moore, A. M., & Geary, D. C. (2016). Countries with higher levels of gender equality show larger national sex differences in mathematics anxiety and relatively lower parental mathematics valuation for girls. *Plos One*, *11*, e0153857. http://doi.org/10.1371/journal.pone.0153857

Sullivan, J., & Barner, D. (2014). The development of structural analogy in number-line estimation. *Journal of Experimental Child Psychology*, *128*, 171–89. http://doi.org/10.1016/j.jecp.2014.07.004

Thien, L. M., & Ong, M. Y. (2015). Malaysian and Singaporean students' affective characteristics and mathematics performance: evidence from PISA 2012. *SpringerPlus*, *4*(1), 563. http://doi.org/10.1186/s40064-015-1358-z

Tournaki, N. (2003). Effect of Student Characteristics on Teachers' Predictions of Student Success. *The Journal of Educational Research*, *96*(5), 310–319. http://doi.org/10.1080/00220670309597643

Trick, L. M., & Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: Evidence for limited capacity preattentive processing. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(2), 331–351. http://doi.org/10.1037/0096-1523.19.2.331

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*(1), 80–102. http://doi.org/10.1037/0033-295X.101.1.80

Turconi, E., Jemel, B., Rossion, B., & Seron, X. (2004). Electrophysiological evidence for differential processing of numerical quantity and order in humans. *Cognitive Brain Research*, *21*(1), 22–38. http://doi.org/10.1016/j.cogbrainres.2004.05.003

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 166–179. http://doi.org/10.1037/0278-7393.18.1.166

van Oeffelen, M. P., & Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Perception & Psychophysics*, *32*(2), 163–170. http://doi.org/10.3758/BF03204275

Van Rinsveld, A., Brunner, M., Landerl, K., Schiltz, C., & Ugen, S. (2015). The relation between language and arithmetic in bilinguals: insights from different stages of language acquisition. *Frontiers in Psychology*, *6*, 265. http://doi.org/10.3389/fpsyg.2015.00265

Vanbinst, K., Ansari, D., Ghesquière, P., & Smedt, B. De. (2016). Symbolic numerical magnitude proceing is as important to arithmetic as phonological awarene is to reading. *PLoS ONE*, *11*(3), 1–11. http://doi.org/10.1371/journal.pone.0151045

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., & Chang, A. (2014). Deconstructing Building Blocks: Preschoolers’ Spatial Assembly Performance Relates to Early Mathematical Skills. *Child Development*, *85*(3), 1062–1076. http://doi.org/10.1111/cdev.12165

Vogel, S. E., Goffin, C., & Ansari, D. (2015). Developmental specialization of the left parietal cortex for the semantic representation of Arabic numerals: An fMR-adaptation study. *Developmental Cognitive Neuroscience*, *12*(1), 61–73. http://doi.org/10.1016/j.dcn.2014.12.001

Vogel, S. E., Remark, A., & Ansari, D. (2014). Differential processing of symbolic numerical magnitude and order in first-grade children. *Journal of Experimental Child Psychology*, *129C*, 26–39. http://doi.org/10.1016/j.jecp.2014.07.010

Vuokko, E., Niemivirta, M., & Helenius, P. (2013). Cortical activation patterns during subitizing and counting. *Brain Research*, *1497*, 40–52. http://doi.org/10.1016/j.brainres.2012.12.019

Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition*, *119*(1), 10–22. http://doi.org/10.1016/j.cognition.2010.11.014

Wang, J., Odic, D., Halberda, J., & Feigenson, L. (2016). Changing the precision of preschoolers’ approximate number system representations changes their symbolic math performance. *Journal of Experimental Child Psychology*, *147*, 82–99. http://doi.org/10.1016/j.jecp.2016.03.002

Wang, Z., Hart, S. A., Kovas, Y., Lukowski, S., Soden, B., Thompson, L. A., … Petrill, S. A. (2014). Who is afraid of math? Two sources of genetic variance for mathematical anxiety. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *55*(9), 1056–64. http://doi.org/10.1111/jcpp.12224

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's Past is Prologue: Relations Between Early Mathematics Knowledge and High School Achievement. *Educational Researcher (Washington, D.C. : 1972)*, *43*(7), 352–360. http://doi.org/10.3102/0013189X14553660

Wong, T. T. Y., Ho, C. S. H., & Tang, J. (2016). The relation between ANS and symbolic arithmetic skills: The mediating role of number-numerosity mappings. *Contemporary Educational Psychology*, *46*, 208–217. http://doi.org/10.1016/j.cedpsych.2016.06.003

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*(2), 155–193. http://doi.org/10.1016/0010-0277(90)90003-3

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*(2), 220–251. http://doi.org/10.1016/0010-0285(92)90008-P

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. http://doi.org/10.1016/S0010-0277(99)00066-9

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science*, *8*(1), 88–101. http://doi.org/10.1111/j.1467-7687.2005.00395.x

Zebian, S. (2005). Linkages between Number Concepts, Spatial Thinking, and Directionality of Writing: The SNARC Effect and the REVERSE SNARC Effect in English and Arabic Monoliterates, Biliterates, and Illiterate Arabic Speakers. *Journal of Cognition and Culture*, *5*(1), 165–190. http://doi.org/https://doi.org/10.1163/1568537054068660

Appendix 1

For each grade participating to the study, we calculated values that quantify aspects of the grade distribution (i.e., skew and kurtosis) and compared the distribution of to a normal by using the Shapiro-Wilk test, *W*. Also, figures from 4.19 to 4.26 show math grade histograms from 3rd to 10th grade.

a) Third Grade (8-9-year-old children)

At the 2nd evaluation, the average math mark was 6.92 (SE = 0.15; range 5 to 9). School math marks were non-normally distributed, $W = 0.86$, $p < .001$; with skewness of 0.25 (*skewness / 2SE* = 0.5; not significant) and Kurtosis of -1.36 (*kurtosis / 2SE* = -1.37; significant at $p < .01$). The negative value of kurtosis indicated a flat and light-tailed distribution.

At 3rd evaluation, the average math mark was 7.36 (SE = 0.15) and ranged from 5 to 10. School math marks at 3rd evaluation were non-normally distributed, $W = 0.87$, $p <.001$; with skewness of 0.23 (skewness / 2SE = 0.45; not significant) and Kurtosis of -1.31 (kurtosis / 2SE = -1.31; significant at $p < .01$). The negative value of kurtosis indicated a flat and light-tailed distribution.

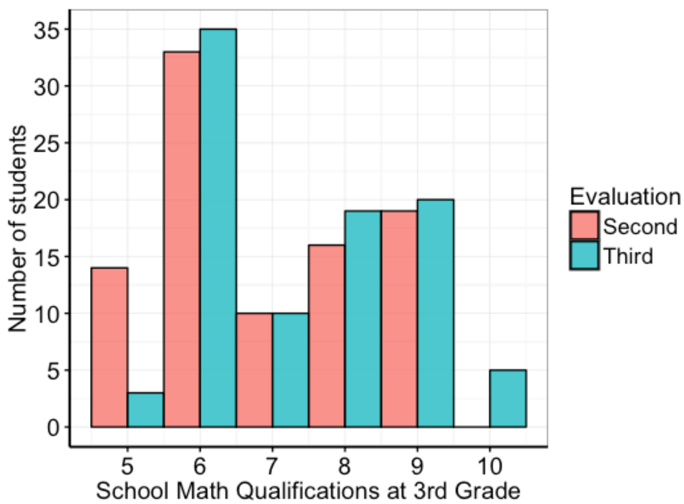


Figure 4.19. *Histogram for School Math Marks at third grade (age= 8-9 year olds; n = 92).*

b) Fourth Grade (9-10-year-old children)

At 2nd evaluation, the average math mark was 8.22 (*SE* = 0.11) and ranged from 5.1 to 9.9. School math marks at 2nd evaluation were non-normally distributed, $W = 0.94$, $p < .001$; with skewness of -0.86 (*skewness / 2SE* = -1.66; significance at $p < .001$) and Kurtosis of 0.91 (*kurtosis / 2SE* = 0.88; not significant). The negative skewness indicated that frequent scores were clustered at the higher end, on the right of the distribution and the tail pointed towards the lower scores.

At 3rd evaluation, the average math mark was 7.94 (*SE* = 0.13) and ranged from 4.1 to 10. School math marks at 3rd evaluation were non-normally distributed, $W = 0.96$, $p < .01$; with skewness of -0.7 (*skewness / 2SE* = -1.34; significance at $p < .001$) and Kurtosis of 0.28 (*kurtosis / 2SE* = 0.27; not significant). The negative skewness indicated that frequent scores were clustered at the higher end, on the right of the distribution and the tail pointed towards the lower scores.

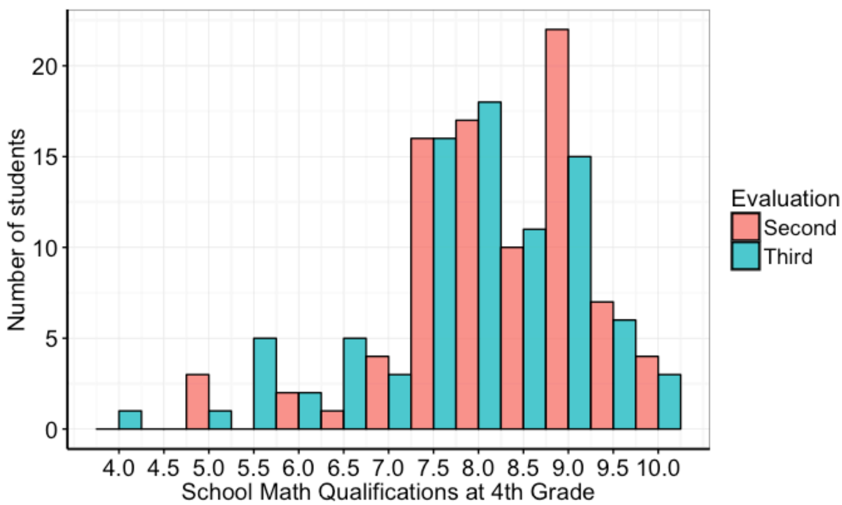


Figure 4.20. *Histogram for School Math Marks at fourth grade (age= 9-10 year olds; n = 86).*

c) Fifth Grade (10-11-year-old children)

At 2nd evaluation, the average math mark was 7.8 (*SE* = 0.2) and ranged from 4 to 10. School math marks at 2nd evaluation were non-normally distributed, $W = 0.94$, $p < .01$; with skewness of -0.52 (*skewness / 2SE* = -0.84; not significant) and Kurtosis of -0.83 (*kurtosis / 2SE* = -0.69; not significant).

At 3rd evaluation, the average math mark was 7.67 (*SE* = 0.18) and ranged from 4.8 to 9.93. School math marks at 3rd evaluation were non-normally distributed, $W = 0.95$, $p < .05$; with skewness of -0.07 (*skewness / 2SE* = -0.12; not significant) and Kurtosis of -1.19 (*kurtosis / 2SE* = -0.99; not significant).

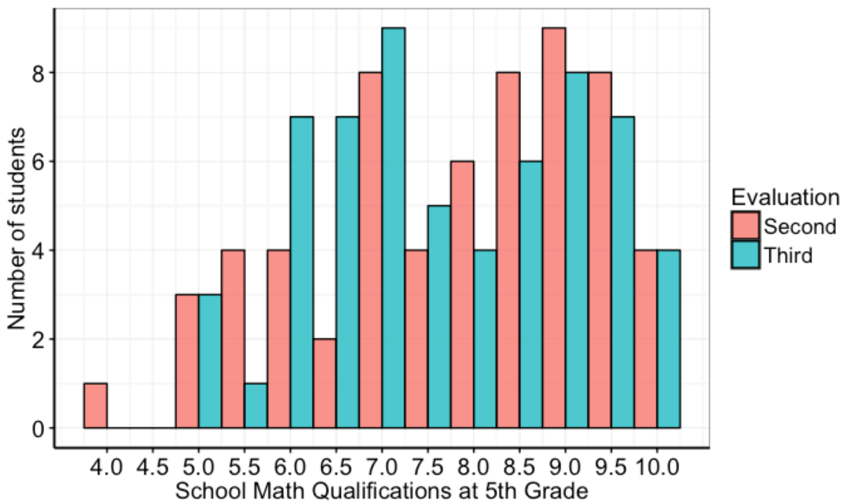


Figure 4.21. *Histogram for School Math Marks at fifth grade (age= 10-11 year olds; n = 61).*

d) Sixth Grade (11-12-year-old students)

At 2nd evaluation, the average math mark was 7.6 (*SE* = 0.16) and ranged from 2.3 to 9.8. School math marks at 2nd evaluation were non-normally distributed, $W = 0.94$, $p < .001$; with skewness of -0.99 (*skewness / 2SE* = -1.81; significance at $p < .001$)

and Kurtosis of 1.64 (*kurtosis / 2SE* = 1.52; significance at $p < .01$). The negative skewness indicated that frequent scores were clustered at the higher end, on the right of the distribution and the tail pointed towards the lower scores. The positive significant value of kurtosis indicated a pointy and heavy-tailed distribution.

At 3rd evaluation, the average math mark was 7.9 (SE = 0.13) and ranged from 3.2 to 9.9. School math marks at 3rd evaluation were non-normally distributed, $W = 0.94$, $p <.001$; with skewness of -1.06 (*skewness / 2SE* = -1.95; significance at $p < .001$) and Kurtosis of 2.15 (*kurtosis / 2SE* = 2; significance at $p < .001$). The negative skewness indicated that frequent scores were clustered at the higher end, on the right of the distribution and the tail pointed towards the lower scores. The positive significant value of kurtosis indicated a pointy and heavy-tailed distribution.

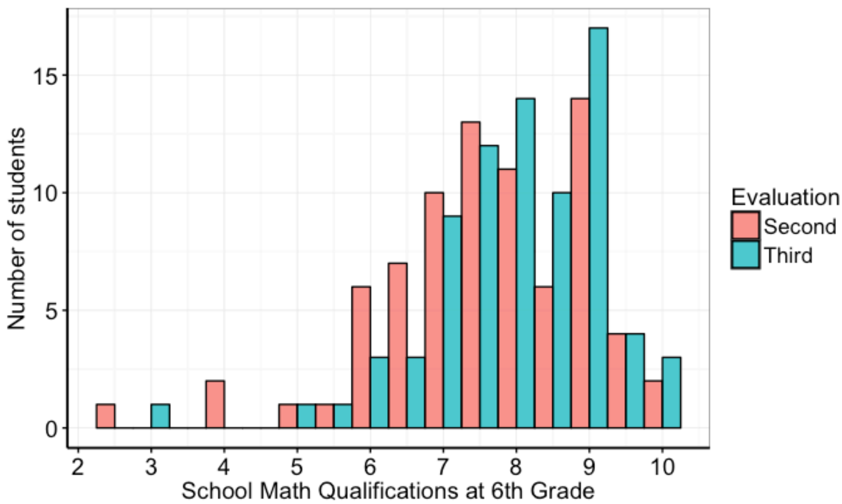


Figure 4.22. *Histogram for School Math Marks at sixth grade (age= 11-12 year olds; n = 78).*

e) Seventh Grade (12-13-year-old students)

At 2nd evaluation, the average math mark was 7 (*SE* = 0.26) and ranged from 1 to 10. School math marks at 2nd evaluation were non-normally distributed, *W* = 0.93, *p* <.001; with skewness of -0.75 (*skewness / 2SE* = -1.32; significance at *p* < .01) and Kurtosis of -0.08 (*kurtosis / 2SE* = -0.07; not significant). The negative skewness indicated that frequent scores were clustered at the higher end, on the right of the distribution and the tail pointed towards the lower scores.

At 3rd evaluation, the average math mark was 7.2 (*SE* = 0.26) and ranged from 1.6 to 10. School math marks at 3rd evaluation were non-normally distributed, *W* = 0.92, *p* <.001; with skewness of -0.75 (*skewness / 2SE* = -1.32; significance at *p* < .01) and Kurtosis of -0.24 (*kurtosis / 2SE* = -0.21; not significant). The negative skewness indicated that frequent scores were clustered at the higher end, on the right of the distribution and the tail pointed towards the lower scores.

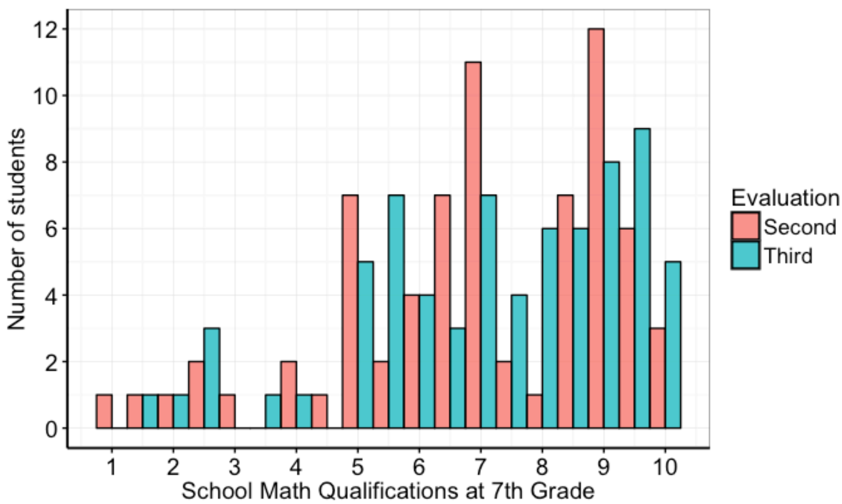


Figure 4.23. *Histogram for School Math Marks at seventh grade (age= 12-13 year olds; n = 71).*

f) Eighth Grade (13-14-year-old students)

At 2nd evaluation, the average math mark was 6.7 (SE = 0.25) and ranged from 2.4 to 9.6. School math marks at 2nd evaluation were non-normally distributed, W = 0.95, p =.03; with skewness of -0.38 (*skewness / 2SE* = -0.6; not significant) and Kurtosis of -0.87 (*kurtosis / 2SE* = -0.7; not significant).

At 3rd evaluation, the average math mark was 7.1 (SE = 0.25) and ranged from 1.5 to 9.6. School math marks at 3rd evaluation were non-normally distributed, W = 0.91, p <.001; with skewness of -1.07 (*skewness / 2SE* = -1.7; significance at p < .001) and Kurtosis of 0.9 (*kurtosis / 2SE* = 0.73; not significant). The negative skewness indicated that frequent scores were clustered at the higher end, on the right of the distribution and the tail pointed towards the lower scores.

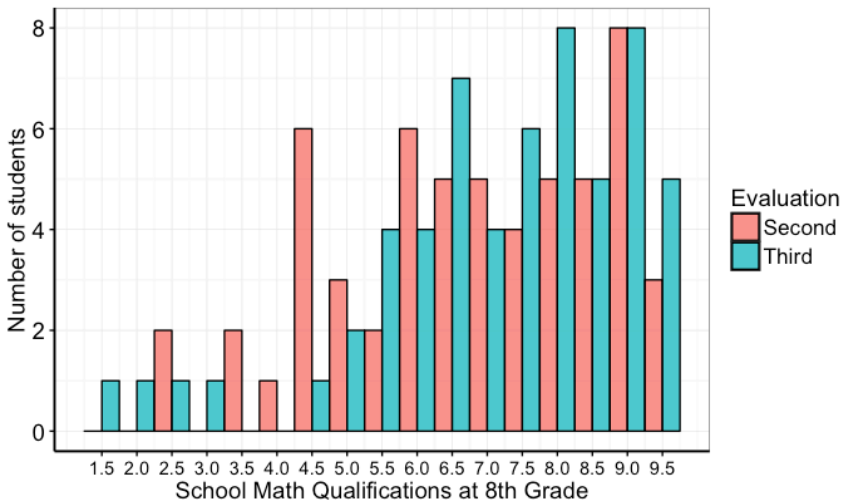


Figure 4.24. *Histogram for School Math Marks at eighth grade (age= 13-14 year olds; n = 58).*

g) Ninth Grade (14-15-year-old students)

At 2nd evaluation, the average math mark was 7.1 (*SE* = 0.27) and ranged from 4 to 10. School math marks at 2nd evaluation were non-normally distributed, $W = 0.92$, $p < .01$; with skewness of -0.49 (*skewness / 2SE* = -0.66; not significant) and Kurtosis of -0.74 (*kurtosis / 2SE* = -0.51; not significant).

At 3rd evaluation, the average math mark was 7.4 (*SE* = 0.29) and ranged from 3.5 to 9.9. School math marks at 3rd evaluation were non-normally distributed, $W = 0.93$, $p < .05$; with skewness of -0.31 (*skewness / 2SE* = -0.42; not significant) and Kurtosis of -1.22 (*kurtosis / 2SE* = -0.83; not significant).

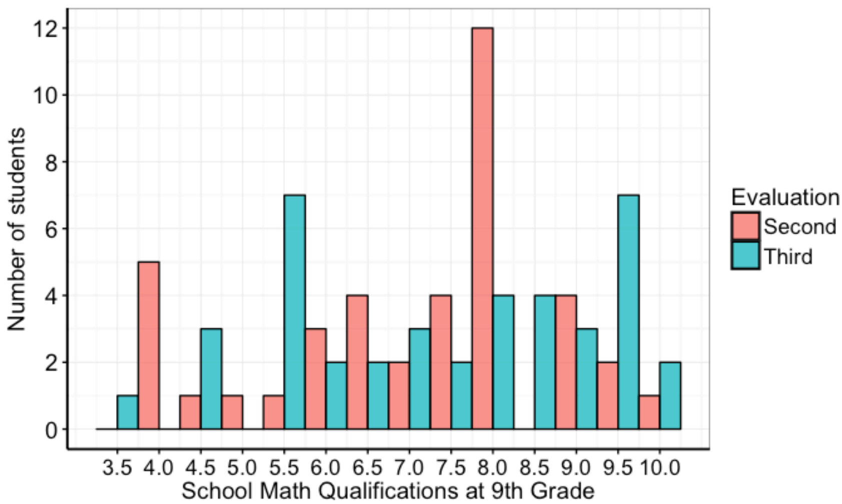


Figure 4.25. *Histogram for School Math Marks at ninth grade (age= 14-15 year olds; n = 40).*

h) Tenth Grade (15-16-year-old students)

At 2nd evaluation, the average math mark was 7.6 (*SE* = 0.29) and ranged from 3 to 10. School math marks at 2nd evaluation were non-normally distributed, $W = 0.94$, $p < .05$; with skewness of -0.63 (*skewness / 2SE* = -0.83; not significant) and Kurtosis of -0.34 (*kurtosis / 2SE* = -0.23; not significant).

At 3rd evaluation, the average math mark was 7.6 (SE = 0.28) and ranged from 3.6 to 10. School math marks at 3rd evaluation were non-normally distributed, W = 0.94, p <.05; with skewness of -0.38 (*skewness / 2SE* = -0.50; not significant) and Kurtosis of -0.83 (*kurtosis / 2SE* = -0.56; not significant).

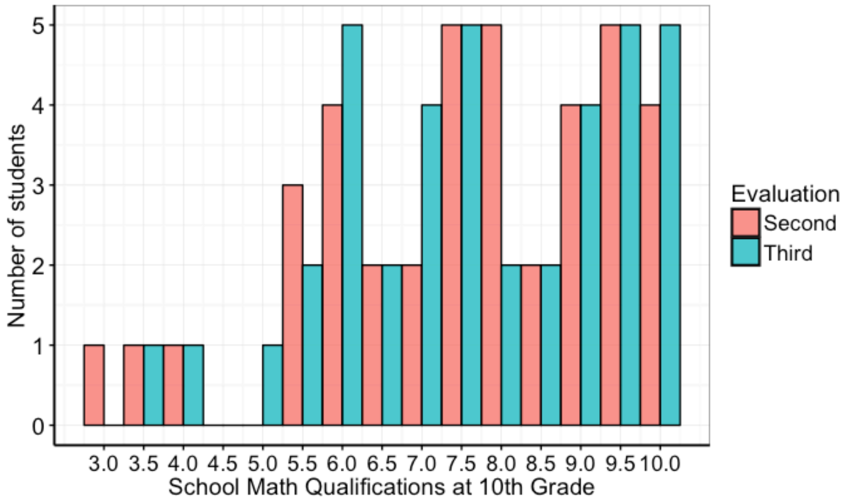


Figure 4.26. *Histogram for School Math Marks at tenth grade (age= 15-16 year olds; n = 39).*

.

Appendix 2

Below, we report detailed analyses, at each age, of the correlations between our tests and school math performance, and of the correlations between the measures of the tests. We invite the reader to refer to section 4.3.3 in the main text for their discussion.

a) Third Grade (8-9-year-old children)

At this age, a total of 92 children could perform the Numerical Estimation task. The percentage of correct answers was significantly related to school math marks at the 2nd and 3rd evaluations, (respectively, $r_s = .32$, $p = .002$; and $r_s = .30$, $p = .004$). Response time was not related to school math marks.

By contrast, there was no correlation between performance in the Quantity Discrimination task (n= 71), both for PCs and RTs, and school math marks.

Because only 71 children of 3rd grade (n= 92) could perform the QD task, we wanted to be sure that the relation between school math performance and NE task that we found was not due to the sample size. Thus, we reduced the NE task sample to the same seventy-one children of the QD task, and we calculated again the correlation. The percentage of correct answers in NE task was still significantly related to school math marks at the 2nd and 3rd evaluations, (respectively, $r_s = .30$, $p < .05$; and $r_s = .24$, $p < .05$; Table 1). Response time was not related to school math marks.

Likewise, to analyze the correlations between the tasks, we proceeded to consider only those participants who performed both tasks (n=71; Table 1).

Concerning the relation between the two tasks, their response times were correlated ($r_s = .39$, $p < .001$): the speed of responses improved with similar profiles across the tasks. However, there was no correlation in the percentage of correct answers of the two tasks. This suggests that the resources employed in one task are not necessarily the same in the other task.

Considering both tasks separately, the correlation between the percentage correct and the response time was significant and positive in both tasks (NE task: $r_s = .37$, $p < .01$; QD task: $r_s = .65$, $p < .0001$); that is, the longer the time to respond was, the higher the percentage of correct answers was.

Also, the response time in the Numerical Estimation task was positive correlated with the percentage of correct responses for both tasks, NE and QD.

Table 1 | Spearman's correlation (r_s) between School Math marks (at 2nd and 3rd Evaluations) and PC (Percent Correct) and RT (Response Time) at NE (Numerical Estimation Task) and QD (Quantity Discrimination task), n=71

	PC NE	RT NE	PC QD	RT QD	2nd Ev	3rd Ev
PC NE	1	$.37^{**}$	.19	.14	$.30^{*}$	$.24^{*}$
RT NE	$.37^{**}$	1	$.29^{*}$	$.39^{***}$	.01	.11
PC QD	.19	$.29^{*}$	1	$.65^{****}$	-.06	.07
RT QD	.14	$.39^{***}$	$.65^{****}$	1	-.22	-.07
2nd Ev	$.30^{*}$	.01	-.06	-.22	1	$.83^{****}$
3rd Ev	$.24^{*}$	.11	.07	-.07	.83	1

ns = not significant ($p > .05$), $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$, $^{****}p < .0001$

b) Fourth Grade (9-10-year-old children)

We restricted our analysis to those children that performed both tasks (n= 85), the Numerical Estimation (n=85) and the Quantity Discrimination task (n=86; Table 2).

The Percentage of Correct answers at Numerical Estimation task (PC NE), was correlated to school math marks, both at the 2nd and 3rd evaluations (respectively, $r_s = .25$, $p = .02$; and, $r_s = .32$, $p = .003$). By contrast, response time was not related to school math marks.

For the Quantity Discrimination task, the percentage of correct answers (PC QD) only correlated with school marks the period of the 3rd evaluation, ($r_s = .32$, $p = .003$). No correlation occurred with response time.

Concerning the relation between the two tasks, response time correlated positively, ($r_s = .32$, $p < .01$). In addition, also the percentage of correct responses between the tasks correlated (r_s = .25, p < .05), unlike what we found at third grade.

Considering both tasks separately, PC and RT had a positive and highly significant correlation in QD task (r_s = .6, p < .0001) and a significant correlation in NE task (r_s = .34, p < .01). Thus, the longer the response time was, the higher the percentage of correct answers was in both tasks.

Table 2 | Spearman's correlation (r_s) between School Math marks (at 2nd and 3rd Evaluations) and PC (Percent Correct) and RT (Response Time) at NE (Numerical Estimation Task) and QD (Quantity Discrimination task), n=85

	PC NE	RT NE	PC QD	RT QD	2nd Ev	3rd Ev
PC NE	1	.34**	.25*	.08	.25*	.32**
RT NE	.34**	1	.12	.32**	.11	.07
PC QD	.25*	.12	1	.60****	.18	.31**
RT QD	.08	.32**	.60****	1	.12	.12
2nd Ev	.25*	.11	.18	.12	1	.70****
3rd Ev	.32**	.07	.31**	.12	.70****	1

ns = not significant (p > .05), *p < .05, **p < .01, ***p < .001, ****p < .0001

c) Fifth Grade (10-11-year-old children)

All 61 participants performed both tasks. Thus the analysis of correlations could be calculated with all the cohort (Table 3).

The percentage of correct answers at the Numerical Estimation task (PC NE) strongly correlated with the school math marks at the 2nd and 3rd evaluations (r_s = .53, p < .0001 and r_s = .51, p < .0001, respectively). Response time at Numerical Estimation task (RT NE) tended to correlate with school math marks at the 2nd evaluation (r_s = .25, p =.052) and was more clearly related to it at the 3rd evaluation (r_s = .31, p < .05).

There was no correlation between the percentage of correct responses at the Quantity Discrimination task and school math marks. Response time RT QD negatively correlated with school math marks at the 2nd evaluation, (r_s = -.29, p < .05) and tended to do so at the 3rd evaluation (r_s = -.23, p =.07).

Concerning the relation between tasks, the percentage of correct answers in one task correlated with that of the other task (r_s = .35, p < .01). Also, there was a positive correlation between the response time of the two tasks (r_s = .20, p < .05). These data suggest that the tasks could be related at this age.

Considering both tasks separately, the correlation between the percentage correct and the response time was significant and positive in both tasks (NE task: r_s = .49, p < .0001; QD task: r_s = .58, p < .0001). That is, the longer the time to respond was, the higher the percentage of correct answers was.

Table 3 | Spearman's correlation (r_s) between School Math marks (at 2nd and 3rd Evaluations) and PC (Percent Correct) and RT (Response Time) at NE (Numerical Estimation Task) and QD (Quantity Discrimination task), n=61

	PC NE	RT NE	PC QD	RT QD	2nd Ev	3rd Ev
PC NE	1	$.49^{****}$	$.35^{**}$	-.22	$.53^{****}$	$.51^{****}$
RT NE	$.49^{****}$	1	$.43^{***}$	$.28^{*}$	.25	$.31^{*}$
PC QD	$.35^{**}$	$.43^{***}$	1	$.58^{****}$	.12	.11
RT QD	-.22	$.28^{*}$	$.58^{****}$	1	$-.29^{*}$	-.23
2nd Ev	$.53^{****}$	.25	.12	$-.29^{*}$	1	$.79^{****}$
3rd Ev	$.51^{****}$	$.31^{*}$	.11	-.23	$.79^{****}$	1

ns = not significant (p > .05), $^{*}p$ < .05, $^{**}p$ < .01, $^{***}p$ < .001, $^{****}p$ < .0001

d) Sixth Grade (11-12-year-old students)

We restricted our analysis to those children that performed both the Numerical Estimation (n=78), and the Quantity Discrimination task (n=77; Table 4).

The percentage of correct answers at Numerical Estimation task (PC NE), correlated with school math marks both at the 2nd and the 3rd evaluation (r_s = .34, p = .002; and, r_s = .44, p < .0001). Response time in this task (RT NE) did not correlate to school math marks.

As for the Quantity Discrimination task, PC correlated to school math marks, both at the 2nd and the 3rd evaluation (r_s = .37, p < .01; r_s = .45, p < .0001) and so did the RT at both evaluations ((r_s = .31, p < .01; r_s = .36, p < .01).

Concerning the relation between tasks, the percentage of correct answers correlated (r_s = .36, p < .05).

Considering both tasks separately, PC and RT had a positive and highly significant correlation in both tasks (NE task: r_s = .40, p < .001; QD task: r_s = .59, p < .0001). That is, the longer the time to respond was, the higher the percentage of correct answers was, in either task.

Table 4 | Spearman's correlation (r_s) between School Math marks (at 2nd and 3rd Evaluations) and PC (Percent Correct) and RT (Response Time) at NE (Numerical Estimation Task) and QD (Quantity Discrimination task), n=77

	PC NE	RT NE	PC QD	RT QD	2nd Ev	3rd Ev
PC NE	1	.40***	.36**	.13	.34**	.44****
RT NE	.40***	1	.22	.16	.04	.15
PC QD	.36**	.22	1	.59****	.37**	.45****
RT QD	.13	.16	.59****	1	.31**	.36**
2nd Ev	.34**	.04	.37**	.31**	1	.78****
3rd Ev	.44****	.15	.45****	.36**	.78****	1

ns = not significant (p > .05), $^{*}p$ < .05, $^{**}p$ < .01, $^{***}p$ < .001, $^{****}p$ < .0001

e) Seventh Grade (12-13-year-old students)

All 71 participants performed both tasks. Thus, the analysis of correlations could be calculated with all the cohort (Table 5).

The percentage of correct answers in the Numerical Estimation task correlated to school math marks both at the 2nd and 3rd evaluation, (r_s = .31, p = .008; r_s = .40, p = .0006). Response time was not related to school math marks.

The percentage of correct answers in the Quantity Discrimination task (PC QD) also correlated with school math marks in both test evaluations (r_s = .37, p = .0014; and r_s = .44, p = .0001). Response time (RT QD) positively correlated with school math marks in the 3rd evaluation (r_s = .26, p < .05).

Concerning the relation between tasks, the percentage of correct answers (PC NE and PC QD) correlated (r_s = .39, p < .001), as well as the response times (RT NE and RT DQ; r_s = .37, p < .01).

These data suggest a strong relation between the two tasks at this age.

Considering both tasks separately, PC and RT highly correlated in both tasks (NE task: r_s = .41, p < .001; QD task: r_s = .62, p < .0001); that is, the greater the time to respond was, the higher the percentage of correct answers was.

Table 5 | Spearman's correlation (r_s) between School Math marks (at 2nd and 3rd Evaluations) and PC (Percent Correct) and RT (Response Time) at NE (Numerical Estimation Task) and QD (Quantity Discrimination task), n=71

	PC NE	RT NE	PC QD	RT QD	2nd Ev	3rd Ev
PC NE	1	$.41^{***}$	$.39^{***}$	.22	$.31^{**}$	$.40^{***}$
RT NE	$.41^{***}$	1	$.29^{*}$	$.37^{**}$	.10	.13
PC QD	$.39^{***}$	$.29^{*}$	1	$.62^{****}$	$.37^{**}$	$.44^{***}$
RT QD	.22	$.37^{**}$	$.62^{****}$	1	.21	$.26^{*}$
2nd Ev	$.31^{**}$	.10	$.37^{**}$	.21	1	$.86^{****}$
3rd Ev	$.40^{***}$	.13	$.44^{***}$	$.26^{*}$	$.86^{****}$	1

ns = not significant (p > .05), $^{*}p$ < .05, $^{**}p$ < .01, $^{***}p$ < .001, $^{****}p$ < .0001

f) Eighth Grade (13-14-year-old students)

All 58 participants performed both tasks. Thus, the analysis of correlations could be calculated with all the cohort (Table 6).

The performance in the Numerical Estimation task (PC and RT) was not related to the school math marks.

In the Quantity Discrimination task, the percentage of correct answers correlated with the school math marks of the 2nd evaluation (r_s = .26, p < .05), but not the third evaluation. There was no correlation between response time and the school evaluations.

Concerning the relation between tasks, the percentage of correct answers in one task correlated with the percentage of correct answer in the other task (PC NE and PC QD; r_s = .41, p < .01). Also, the response time in the two tasks positively correlated (r_s = .48, p < .001). Overall, the data suggests that the tasks could be related at this age.

Considering both tasks separately, in both cases PC and RT were positively and strongly correlated (NE task: r_s = .59, p < .0001; QD task: r_s = .56, p < .0001); again, the longer the response time was, the higher the percentage of correctness was, in both tasks.

Table 6 | Spearman's correlation (coefficient r_s) between School Math marks (at 2nd and 3rd Evaluations) and PC (Percent Correct) and RT (Response Time) at NE (Numerical Estimation Task) and QD (Quantity Discrimination task), n=58

	PC NE	RT NE	PC QD	RT QD	2nd Ev	3rd Ev
PC NE	1	.59****	.41**	.20	.13	.17
RT NE	.59****	1	.34**	.48***	.02	.05
PC QD	.41**	.34**	1	.56****	.26*	.18
RT QD	.20	.48***	.56****	1	-.02	-.08
2nd Ev	.13	.02	.26*	-.02	1	.82****
3rd Ev	.17	.05	.18	-.08	.82****	1

ns = not significant (p > .05), *p < .05, **p < .01, ***p < .001, ****p < .0001

f) Nineth Grade (14-15-year-old students)

All 40 participants performed both tasks. Thus, the analysis of correlations could be calculated with all the cohort (Table 7).

In this age group, we found no relation between either measure of performance in the Numerical Estimation task (percentage correct or speed of responses), and students' school math marks.

For the Quantity Discrimination task, the percentage of correct answers did not correlate with school math marks. Response time positively correlated with school math marks in both evaluations (r_s = .34, p < .05; r_s = .42, p < .01), indicating that pupils who responded more slowly also had higher marks.

Concerning the relation between tasks, no relation was found in any of the two measures of performance, unlike what we found in the previous age groups.

Considering both tasks separately, only for the QD task, the percentage of correct answers and response time were positively correlated ($r_s = .45$, $p < .01$).

Table 7 | Spearman's correlation (coefficient r_s) between School Math marks (at 2nd and 3rd Evaluations) and PC (Percent Correct) and RT (Response Time) at NE (Numerical Estimation Task) and QD (Quantity Discrimination task), n=40

	PC NE	RT NE	PC QD	RT QD	2nd Ev	3rd Ev
PC NE	1	.21	.11	.02	.21	.24
RT NE	.21	1	-.01	.25	.09	.03
PC QD	.11	-.01	1	$.45^{**}$	.24	.20
RT QD	.02	.25	$.45^{**}$	1	$.34^{*}$	$.42^{**}$
2nd Ev	.21	.09	.24	$.34^{*}$	1	$.79^{****}$
3rd Ev	.24	.03	.20	$.42^{**}$	$.79^{****}$	1

ns = not significant ($p > .05$), $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$, $^{****}p < .0001$

f) Tenth Grade (15-16-year-old students)

All 39 participants performed both tasks. Thus, the analysis of correlations could be calculated with all the cohort (Table 8).

In this age group, we found no relation between either measure of performance in the Numerical Estimation task (percentage correct or speed of responses), and students' school math marks.

For the Quantity Discrimination task, the percentage of correct answers correlated to school math marks both for the 2nd and the 3rd evaluation ($r_s = .42$, $p < .01$ and $r_s = .39$, p <.05), but there was no correlation between speed of responses and school math marks.

Analyzing the relation between tasks, there was not any significant correlation, although the response times of the two tasks (RT NE and RT DQ), tend to be positively related ($r_s = .30$, $p = 0.07$). Thus, the lack of correlation between both tasks was similar to what we found in the previous age group, the 9th grade; and different to the rest of the grades.

Considering both tasks separately, in both cases PC and RT were positively correlated (NE task: r_s = .35, p < .05; QD task: r_s = .51, p < .001); indicating that slower response time gave more correct responses, in both tasks.

For the Numerical Estimations task, response time was also positive correlated with the percentage correct of the Quantity Discriminations task (r_s = .33, p < .5). Therefore, at 15-16 year olds, being quicker at responding did not mean doing it better, rather to the contrary.

Table 8 | Spearman's correlation (coefficient r_s) between School Math marks (at 2nd and 3rd Evaluations) and PC (Percent Correct) and RT (Response Time) at NE (Numerical Estimation Task) and QD (Quantity Discrimination task), n=39

	PC NE	RT NE	PC QD	RT QD	2nd Ev	3rd Ev
PC NE	1	.35*	.07	-.24	-.08	-.08
RT NE	.35*	1	.33*	.30	.07	.05
PC QD	.07	.33	1	.51***	.42**	.39*
RT QD	-.24	.30	.51***	1	.17	.19
2nd Ev	-.08	.07	.42**	.17	1	.92****
3rd Ev	-.08	.05	.39*	.19	.92****	1

ns = not significant (p > .05), $^{*}p$ < .05, $^{**}p$ < .01, $^{***}p$ < .001, $^{****}p$ < .0001