Learning and Instruction 50 (2017) 1–13

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Nonsymbolic and symbolic magnitude comparison skills as longitudinal predictors of mathematical achievement

Iro Xenidou-Dervou [a, b, *], Dylan Molenaar [d], Daniel Ansari [c], Menno van der Schoot [b], Ernest C.D.M. van Lieshout [b]

[a] Mathematics Education Centre, Loughborough University, Loughborough, UK
[b] Section Educational Neuroscience, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[c] Numerical Cognition Laboratory, Department of Psychology and Brain and Mind Institute, University of Western Ontario, London, Ontario, Canada
[d] Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands

A R T I C L E I N F O

Article history:
Received 17 June 2015
Received in revised form
16 September 2016
Accepted 6 November 2016
Available online 28 November 2016

Keywords:
Cognitive development
Mathematical cognition
Nonsymbolic magnitude comparison
Symbolic magnitude comparison
Approximate number system

A B S T R A C T

What developmental roles do nonsymbolic (e.g., dot arrays) and symbolic (i.e., Arabic numerals) magnitude comparison skills play in children's mathematics? We assessed a large sample in kindergarten, grade 1 and 2 on two well-known nonsymbolic and symbolic magnitude comparison measures. We also assessed children's initial IQ and developing Working Memory (WM) capacities. Results demonstrated that symbolic and nonsymbolic comparison had different developmental trajectories; the first underwent larger developmental improvements. Both skills were longitudinal predictors of children's future mathematical achievement above and beyond IQ and WM. Nonsymbolic comparison was moderately predictive only in kindergarten. Symbolic comparison, however, was a robust and consistent predictor of future mathematics across all three years. It was a stronger predictor compared to nonsymbolic, and its predictive power at the early stages was even comparable to that of IQ. Furthermore, the present results raise several methodological implications regarding the role of different types of magnitude comparison measures.

1. Introduction

The question of what underlies the development of mathematical achievement has attracted a lot of attention the last decades. The reason is simple: mathematical skills play a prominent role in our cognitive development and life success (e.g., Dougherty, 2003; Reyna & Brainerd, 2007). Numbers are everywhere and they can take many forms: for example, there is the *nonsymbolic* representation consisting of five dots on a screen and the *symbolic* representation of the number “5” in its Arabic form. What both of these representations have in common is the “fiveness” of the numerosities' magnitude. Extensive focus has been placed on the early markers of numerical cognition, particularly on the role that nonsymbolic and symbolic magnitude comparison skills play as building blocks of numerical cognition (for reviews see De Smedt, Noël, Gilmore, & Ansari, 2013; Feigenson, Libertus, & Halberda, 2013). Findings so far have been contradictory, and in the literature one notices three striking gaps: a) There is a shortage of longitudinal developmental studies examining whether and how the different magnitude processing *predictors' power dynamically changes* from one grade to another (De Smedt et al., 2013; Noël & Rousselle, 2011). b) *Tasks* with fundamentally different design characteristics and number ranges, have been used interchangeably (De Smedt et al., 2013; Gilmore, Attridge, De Smedt, & Inglis, 2014). c) Domain-general capacities such as *working memory* resources and IQ are rarely controlled for (Hornung, Schiltz, Brunner, & Martin, 2014; Xenidou-Dervou, van Lieshout, & van der Schoot, 2014). The present study strived to fill in these gaps and thereby resolve the existing conflicting findings.

1.1. Nonsymbolic and symbolic magnitude processing

Research has indicated that human and non-human primates may be born with an ability to estimate and manipulate abstract quantities in nature. The Approximate Number System (ANS;

* Corresponding author. Mathematics Education Centre, School of Science, Loughborough University, Loughborough, LE11 3TU, UK.
E-mail address: I.Xenidou-Dervou@lboro.ac.uk (I. Xenidou-Dervou).

http://dx.doi.org/10.1016/j.learninstruc.2016.11.001
0959-4752/

Dehaene, 2011) is thought to be a pre-linguistic cognitive system where magnitudes are represented and processed. The ANS enables humans to compare and manipulate nonsymbolic numerosities already from infancy onwards (for reviews see Dehaene, 2011; Feigenson et al., 2013; De Smedt et al., 2013). Of course, as humans we also develop higher-order mathematical skills with symbols. So, how does this "innate" ability affect the development of our symbolic processing and what predicts the development of mathematical achievement, nonsymbolic, symbolic processing or both? These questions have generated intense scientific debate since they have important theoretical as well as educational implications (e.g., De Smedt et al., 2013; Noël & Rousselle, 2011). Establishing which early cognitive predictors play an important role, when and how, in the development of mathematics achievement, can inform educational practice, curricula contents and guide early intervention designs (De Smedt et al., 2013). For example, should educational practice focus on training children's nonsymbolic or symbolic skills or perhaps place different focus at different ages?

Some studies suggest that symbolic representations of number directly map onto ones readily accessible nonsymbolic representations, i.e., the ANS (e.g., Lipton & Spelke, 2005; Piazza & Izard, 2009). In this respect, the ANS is viewed as the cognitive foundation that fosters and enhances the development of general mathematics achievement. This has been a compelling theory and several studies have demonstrated relations between ANS measures and general mathematics achievement (Gilmore, McCarthy, & Spelke, 2010; Inglis, Attridge, Batchelor, & Gilmore, 2011; Libertus, Feigenson, & Halberda, 2011; Starr, Libertus, & Brannon, 2013; for a review see; Feigenson et al., 2013). At the same time, however, many studies have failed to find such relations between the ANS and symbolic processing or mathematics achievement (e.g., Bartelet, Vaessen, Blomert, & Ansari, 2014; Holloway & Ansari, 2009; Lyons, Price, Vaessen, Blomert, & Ansari, 2014; Sasanguie, Defever, Maertens, & Reynvoet, 2014; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013). The latter findings seem to suggest that symbolic numbers are processed and acquire meaning in a fundamentally different way (e.g., Lyons, Ansari, & Beilock, 2012). Within this framework, symbolic magnitude processing is viewed as the best predictor of mathematical achievement, not the ANS (De Smedt et al., 2013; Lyons et al., 2014). Perhaps the predominance of symbolic processing as a predictor of children's individual differences in mathematical achievement, may reflect the fact that children may differ in their ability to access the number magnitude of symbols, rather than processing numerosity in itself (Rouselle & Noël, 2007). Nevertheless, if symbolic processing does not directly map one-to-one onto ones pre-existing nonsymbolic representations, then we may expect them to demonstrate different developmental growth rates (Matejko & Ansari, 2016). As an assumed innate ability, nonsymbolic processing is expected to demonstrate less developmental growth compared to symbolic processing, given that the latter focuses on children assessing the magnitude of Arabic digits, and school mathematics instruction primarily teaches children to use digits to conduct basic arithmetic.

As various contradicting results come forth, the predictive roles of nonsymbolic and symbolic magnitude processing across development remain unclear. In a recent review of findings concerning the relationship between mathematics achievement and nonsymbolic and symbolic magnitude processing, De Smedt et al. (2013) acknowledged two factors that may give rise to the patchwork of contradictory results that characterizes the extant literature: a) The age of the participants assessed, and b) The measures used to assess magnitude comparison.

1.2. Inconsistent findings: Possible sources

1.2.1. Age of participants

In order to identify the role that nonsymbolic and symbolic magnitude skills play, longitudinal and developmental studies are clearly necessary. ANS acuity (Halberda & Feigenson, 2008) and symbolic magnitude precision have been shown to increase with age (Holloway & Ansari, 2009; Sasanguie, De Smedt, Defever, & Reynvoet, 2011). Several longitudinal studies have demonstrated ANS acuity before the start of formal school instruction to correlate with or be predictive of later mathematics achievement (Gilmore et al., 2010; Libertus et al., 2011; Mazzocco, Feigenson, & Halberda, 2011; Starr, Libertus, & Brannon, 2013). Furthermore, Inglis et al. (2011) found that ANS acuity correlates with mathematical achievement in childhood but not in adulthood.

These studies, however, did not assess symbolic magnitude processing. With cross-sectional designs, Lyons et al. (2014) and Sasanguie et al. (2013) assessed various nonsymbolic and symbolic measures simultaneously across *primary school children* and found no evidence for nonsymbolic magnitude processing predicting unique variance in children's arithmetic abilities. Instead, only symbolic magnitude processing played a unique role. On the basis of these findings, we expected that the ANS, as a readily accessible system, may play a unique role primarily in kindergarten, before formal mathematics instruction starts (e.g., Gilmore et al., 2010). From grade 1 and onwards, however, the predictive role of symbolic processing would take over (Lyons et al., 2014; Sasanguie et al., 2013). Thus, we hypothesized that the predictive roles of nonsymbolic and symbolic magnitude comparison skills would dynamically change over time. To our knowledge, this is the first study, which – due to its longitudinal design – allowed the examination of whether and how the predictive roles of magnitude comparison skills change across grades.

In contrast to our aforementioned hypothesis, however, Bartelet et al. (2014) demonstrated that in kindergarten only symbolic magnitude skills predicted children's grade 1 mathematics above and beyond nonsymbolic skills. Notably, though, in this study, children's WM capacities were not controlled for. Also, the measures used in this study differed on several aspects from certain other kindergarten studies; for example, the (non)symbolic stimuli were presented simultaneously, not sequentially (e.g. Gilmore et al., 2010; Xenidou-Dervou et al., 2014). In general, one notable difference across the various studies conducted so far is the measures used to assess nonsymbolic and symbolic magnitude processing skills.

1.2.2. Different magnitude comparison measures

Measures used across the literature can differ both on design characteristics as well as numerosity/number ranges but have nevertheless been used interchangeably. Specifically, in one well-known magnitude comparison measure, the stimuli to be compared (*nonsymbolic or symbolic*) are presented *simultaneously* (see for example Fig. 1A). This measure usually entails *small numerosities* within the range of 1 up to 9 (e.g., De Smedt, Verschaffel, & Ghesquière, 2009; Holloway & Ansari, 2008, 2009; Sasanguie et al., 2013; Sasanguie, Van den Bussche, & Reynvoet, 2012). In contrast, another well-known magnitude comparison measure comprises *large numerosities* ranging for example from 6 up to 70. Also, this measure entails several *sequential* steps (see Fig. 1B): the child sees a blue (nonsymbolic or symbolic) numerosity dropping down on the left side of the screen, this is then covered by an occluder, and then a comparison red quantity drops down on the right side of the screen (Barth et al., 2006; Barth, La Mont, Lipton, & Spelke, 2005; De Smedt & Gilmore, 2011;

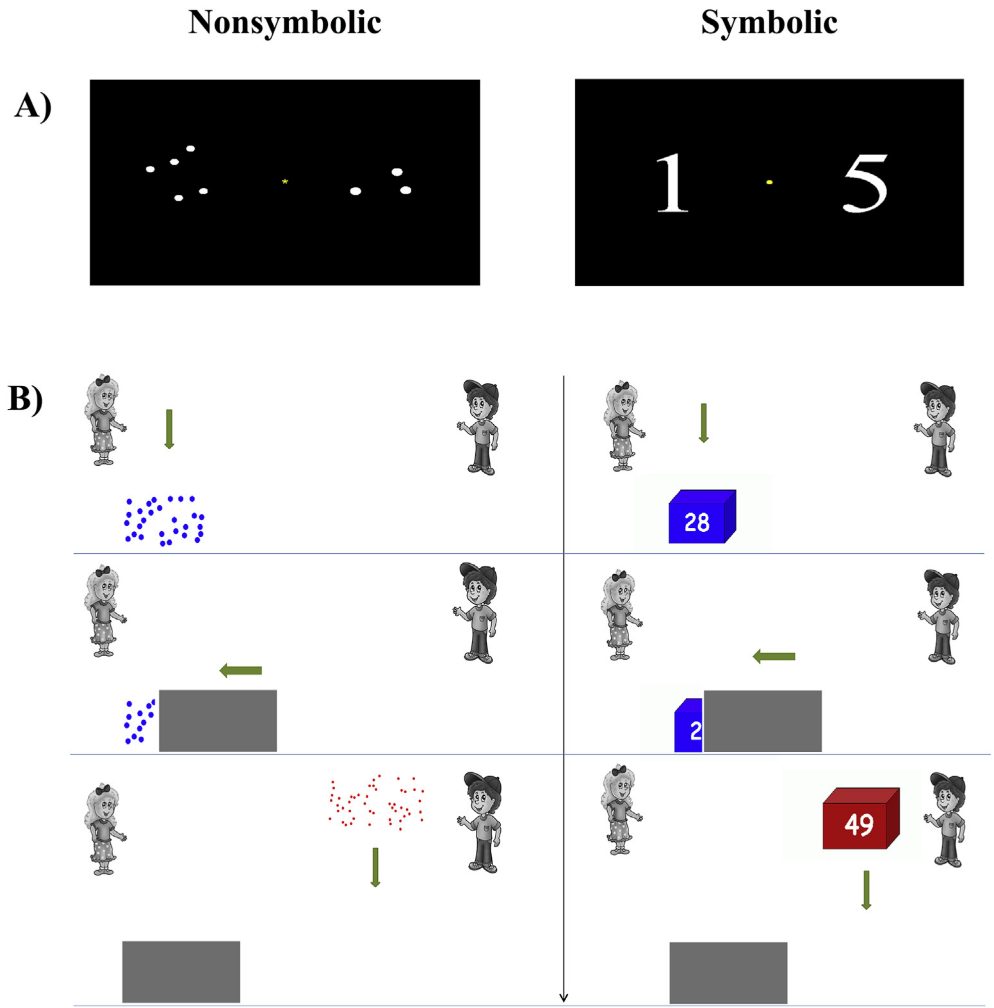

Fig. 1. A) Example trials from the nonsymbolic and symbolic simultaneous-small magnitude comparison tasks. B) Example trials from the nonsymbolic and symbolic sequential-large magnitude comparison tasks.

Gilmore, McCarthy, & Spelke, 2007; Gilmore et al., 2010; Xenidou-Dervou et al., 2013; 2014). Performance across such different measures correlates during childhood (Gilmore et al., 2014) but not in adulthood (Gilmore, Attridge, & Inglis, 2011), which might indicate that they have different developmental trajectories.

Taking a look at these two types of measures (Fig. 1) one may wonder: until what age can accuracy or RT in the symbolic simultaneous-small measure be used to uniquely predict individual differences in mathematical achievement above domain-general capacities and the sequential-large symbolic magnitude measure? Numbers from 1 up to 9 are learned and automatized from early on. We expected to see a ceiling effect in accuracy in the symbolic simultaneous-small measure after formal schooling has started and only its RT data to be predictive of mathematics achievement. It should be noted that it was not within the scope of the present study to examine the difference in the effects of the simultaneous versus sequential presentation-format with either a small or large number-range. We merely assessed the children on the two commonly used different measures of nonsymbolic and symbolic magnitude comparison identified from here forth on the basis of their differential design characteristics, namely “simultaneous-small” and “sequential-large”.

1.2.3. *Accounting for domain-general capacities*

Another important impediment across the existing literature is that most studies do not control for domain-general capacities, such as WM (De Smedt & Gilmore, 2011; Gilmore et al., 2011; Gullick, Sprute, & Temple, 2011). It has recently been demonstrated that nonsymbolic and symbolic magnitude processing call upon different WM resources (Xenidou-Dervou et al., 2014; Xenidou-Dervou, van der Schoot, & van Lieshout, 2015b). WM is a limited-capacity multicomponent cognitive system, which is responsible for the short-term storage and manipulation of information in an online manner when executing cognitive tasks. According to Baddeley's model (Baddeley, 2003, 2012; Repovs & Baddeley, 2006), WM entails the Phonological Loop (PL), which retains phonological information, the Visuospatial Sketchpad (VSSP), which retains visuospatial information and the Central Executive (CE), which monitors, controls and regulates the processes of the other two components and connects them with one's long-term memory. Later, Baddeley added a fourth component to the model: the episodic buffer, a multidimensional passive storage system, which allows elements from the other components to be combined and integrated with long-term memory (Repovs & Baddeley, 2006). As there is little developmental research on this component and the predictors we focused on did not require the integration of information in an episodic manner, the episodic buffer was not addressed.

WM plays a fundamental role in mathematical achievement (DeStefano & LeFevre, 2004; Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, 2013; Raghubar, Barnes, & Hecht, 2010). Recently it has been shown that both nonsymbolic and symbolic magnitude processing are related to WM capacity (Gullick et al., 2011; Hornung, Schiltz, Brunner, & Martin, 2014; Xenidou-Dervou et al., 2014, 2015b). The role of the different WM components and their interactions when conducting a given cognitive task depend upon the characteristics of the task and the age of participants (Friso-van den Bos et al., 2013; Rasmussen & Bisanz, 2005; Simmons, Willis, & Adams, 2012; Xenidou-Dervou et al., 2015b). For nonsymbolic processing it has become evident that WM plays a central role. Specifically, Xenidou-Dervou et al. (2014) demonstrated that children's nonsymbolic processing with sequential steps and large numerosities (6–70) necessitates CE resources. On the other hand, nonsymbolic performance in measures that make use of smaller quantities (1–9) appears to correlate with children's readily accessible VSSP (Xenidou-Dervou et al., 2015b). Especially in the case where the nonsymbolic quantities are presented simultaneously, research suggests that children may primarily rely on the visual characteristics of the stimuli (Gebuis & Reynvoet, 2012; Gilmore et al., 2013). However, when numbers or quantities are presented sequentially, i.e., the first numerosity or number is hidden and participants must remember it in order to compare it with the next numerosity or number, one could argue that more WM load is introduced. Alternatively, one may argue that because the numerosity is hidden in the sequential presentation, it forces the participant to extract a mental numerical representation of this quantity to compare it with the next one – this way the participant may perhaps rely less on the visual features of the dots, constituting it a more effective way of tapping into the ANS. With respect to symbolic processing, there is primarily correlational literature indicating that WM relates with performance in symbolic measures (Gullick et al., 2011; Hornung et al., 2014; Xenidou-Dervou et al., 2015b). In general, however, the role of WM in symbolic magnitude processing is yet unclear.

Overall, it becomes clear that in order to identify the *unique* role that performance on a given magnitude measure plays at a given age, one must control for the effect of WM capacities. For the present study we assessed the participants' performance on a wide range of both math-specific as well as domain-general WM tasks assessed at each developmental stage (kindergarten, grade 1 and grade 2) in order to control for their effects. Another domain-general capacity, which is often not taken into account in the magnitude comparison literature, is that of IQ (De Smedt & Gilmore, 2011). It is no surprise that fluid intelligence is a fundamental predictor of academic performance in general, including mathematics achievement (e.g., Colom, Escorial, Shih, & Privado, 2007; Ritchie, 2015). Given all the recent focus on the predictive power of magnitude comparison skills on mathematical achievement, we examined whether they are predictive of future mathematics achievement even after controlling for children's IQ and if yes, whether their predictive power compares with that of IQ.

1.3. The present study

We aimed to shed light on the conflicting results of studies that focus on nonsymbolic or symbolic magnitude processing as precursors of children's mathematics achievement in different ages. We, therefore, administered the two well-known measures of nonsymbolic and symbolic magnitude comparison in a relatively large Dutch-speaking sample in kindergarten, grade 1 and grade 2. We also assessed these children's IQ in kindergarten and their WM skills at each developmental stage as domain-general control measures. Lastly, their general mathematical achievement was measured at the end of grade 2. In the Netherlands, formal schooling initiates in grade 1, not earlier. Therefore, this sample is unique in the sense that it allows us to examine the developmental transition from kindergarten to formal schooling. Our aim was to address the following two research questions:.

1. Does performance in the two different stimuli formats (nonsymbolic vs. symbolic) and in the two different well-known magnitude comparison types of measures (simultaneous-small vs. sequential-large) develop differently from kindergarten up to grade 2?

Given that formal mathematics instruction (grade 1 and onwards) focuses on the use of symbols and basic arithmetic, we expected that symbolic magnitude processing would demonstrate larger developmental growth rates compared to nonsymbolic magnitude processing, i.e., the two abilities would have different developmental trajectories (Lyons et al., 2012; Matejko & Ansari, 2016; Xenidou-Dervou, Gilmore, van der Schoot, & van Lieshout, 2015a). Furthermore, given the fundamental differences between the two types of measures outlined earlier (simultaneous-small vs. sequential-large), we hypothesized that performance on these measures would also demonstrate different developmental trajectories (Gilmore et al., 2014, 2011).

2. A) Which of the nonsymbolic and symbolic magnitude comparison skills *uniquely* predict future mathematics achievement above and beyond children's IQ and WM capacities at *each year* (kindergarten, grade 1 and grade 2)? B) Do magnitude comparison skills in the subsequent years (after kindergarten) improve the prediction of future mathematics achievement? C) Lastly, which nonsymbolic and symbolic magnitude comparison skills *uniquely* predict future mathematics achievement *across all years* over and above domain-general capacities?

For research questions 2A and 2C, we hypothesized that in kindergarten nonsymbolic (Gilmore et al., 2010; Libertus et al., 2011) and symbolic magnitude processing would uniquely predict future mathematics achievement (Bartelet et al., 2014). However, with the start of formal education (grade 1), symbolic processing would take over (De Smedt et al., 2013; Lyons et al., 2014; Sasanguie, et al., 2013). For 2B, we expected that the dynamic change of the predictive roles of magnitude comparison would improve the prediction of future mathematics achievement; in other words, children's magnitude comparison growth (primarily symbolic) across the years would contribute to the prediction of their future mathematics performance.

2. Method

2.1. Participants

This data is part of a collaborative project known as the MathChild project during which 444 children from 25 schools in the Netherlands were assessed on a number of measures in kindergarten, grade 1 and grade 2. Written consent was acquired from all children's legal guardians. Children identified as extreme outliers, namely those who scored more than three standard deviations above or below the group mean in one or more of the present study's measures were removed from the analyses (40 children). Throughout the three years of assessment, 80 children dropped out primarily due to family relocations. At the last measurement wave (see Table 1), the sample consisted of 326 children (M_{age} = 7.99 years, SD = 0.33, 180 boys, 146 girls). All children spoke Dutch and 96.6% of them held the Dutch nationality. The sample was acquired from middle- to high-SES environments. 33.6% of the children's mothers and 26.2% of their fathers had received middle-level applied education (in the Dutch Educational system: MBO). 42.9% of the mothers and 46.2% of the fathers attended higher levels of education (in the Dutch Educational system: HBO and higher levels).

Table 1
Measurement timeline.

Task	Kindergarten		Grade 1		Grade 2	
	T1	T2	T3	T4	T5	T6
IQ	X					
Dot Matrix, Odd One Out	X		X		X	
Word RF, Word RB		X		X		X
Digit RF, Digit RB		X		X		X
Small-simultaneous	X			X	X	
Large-sequential	X		X		X	
General Math Achievement						X

Note. T1, T3, T5 measurement waves took place in the 1st half of the given academic year (November–December) and T2, T4, T6 in the 2nd half (May–June). The General Math Achievement test (CITO) was administered in June, at the end of grade 2. RF = Recall Forward, RB = Recall Backwards.

2.2. Procedure

All participants were tested individually in quiet settings within their school facilities by trained experimenters with the exception of the IQ and the general mathematics test (CITO). The IQ test was administered in group settings by the experimenters. The CITO ability scores were collected by school staff as part of the usual school tests. The rest of the data of this study comprises a set of tasks administered for the collaborative project across three testing sessions of approximately 20 min in kindergarten and across two sessions of 30 min duration in grade 1 and grade 2. Between two sessions, there was a minimum of a day and a maximum of two weeks. Table 1 depicts the timeline of administration of the materials. All experimenters used the same elaborate protocol with instructions for testing administration across all measurements. Parts of the kindergarten data have been reported in previous studies (Friso-van den Bos et al., 2014; Xenidou-Dervou et al., 2013) as well as the end of grade 2 mathematics achievement data (Cito; Friso-van den Bos, Kroesbergen et al., 2015; Friso-van den Bos, Van Luit et al., 2015). These studies focused on different research questions.

2.3. Materials

All materials, apart from the general mathematics achievement (Cito) and IQ tests, were computerized and presented with E-Prime version 1.2 (Psychological Software Tools, Pittsburgh, PA, USA) in HP Probook 6550b laptops.

2.3.1. Magnitude comparison measures

2.3.1.1. Simultaneous-small. We administered a nonsymbolic and symbolic measure developed on the basis of the widely used "magnitude comparison" measure (Holloway & Ansari, 2009; Sekuler & Mierkiewicz, 1977). These tasks entailed 6 practice and 26 testing trials. During testing, no feedback was provided. In each trial, the child saw two numerosities, one on the right and one on the left side of the screen (see Fig. 1A). Participants were asked to identify which numerosity was larger by pressing the left or the right response box situated in front of them. In a half of the trials, the larger numerosity was presented on the right side of the screen and in the other half, on the left. Children were instructed to respond as correctly and as fast as possible. Numerosities in these tasks ranged from 1 up to 9. The testing trials included all possible numerical pairs with the absolute distances between the comparison numerosities ranging from 1 to 4 (distance 1: 8 trials; distance 2: 7 trials; distance 3: 6 trials; distance 4: 5 trials).

The *nonsymbolic* condition started with an alerting beep sound of 100 ms followed by a 1500 ms warning interval (< >). As depicted in Fig. 1A, the dot stimuli were presented in white on a black background, left and right from a yellow asterisk (fixation point). The response interval lasted until an answer was provided or until a maximum of 5000 ms was reached. To prevent the children from counting the dots, the stimuli were only presented for 840 ms. As in previous studies, continuous quantity variables, were controlled for with the methodology developed by Dehaene, Izard, and Piazza (2005). According to this methodology, dot diameter was constant in half of the trials whereas in the other half, the size of the total dot surface area was constant. Trial order was randomized. For each continuous quantity variable (constant dot size and constant area) and for each numerosity, there was a pool of 16 different dot patterns. The program chose randomly one of these, so that the individual patterns of the dots were randomized as well. Thus, it is assumed that it is unlikely that the responses could be associated with specific dot patterns instead of quantity.

The *symbolic* condition was identical to the nonsymbolic, with the key difference that the corresponding Arabic numeral now replaced the dot stimuli. In this condition, the fixation point was now a dot instead of an asterisk in order to prevent possible confusion with the multiplication sign.

2.3.1.2. Sequential-large. A nonsymbolic and a symbolic version of the commonly used "approximate comparison" measures were used (Barth et al., 2006; Gilmore et al., 2007; Gilmore et al., 2010). These measures included 6 practice and 24 testing trials. Feedback was only provided during practice. The number of practice trials was reduced to two in grade 1 and grade 2, as children were already familiar with this measure.

In the *nonsymbolic* version the children were told that Sarah and Peter receive a set of blue and red dots respectively and were asked to respond to the question "Who got more dots? Sarah or Peter?". Within a trial (Fig. 1B), the following sequence of events took place: 1) An amount of blue dots appeared and dropped on the left side of the screen next to the image of the girl, 2) These were then covered by a grey box, 3) A set of red dots popped up and dropped on the right side of the screen next to the image of the boy. Children were instructed to respond as correctly and as fast as possible by pressing the blue or red response box in front of them. Each animated event lasted 1300 ms and between each event there was a 1200 ms interval. The fast interchange of events prevented counting. The child could respond from the moment the red dots appeared on the screen within a maximum of 7000 ms. Between trials, there was a 300 ms interval. Numerosities ranged from 6 up to 70. The blue array differed from the comparison red array by three ratios: 4:7, 4:6, 4:5 (easy, middle and difficult ratio). There were eight trials for each ratio. In half of the trials the blue array was larger, whereas in the other half the red was larger. Trial order was randomized. To avoid responses being reliant on the physical features of the dots and not quantity per se, dot stimuli followed a commonly used control methodology: Dot size, total dot surface area, total dot contour length and density correlated positively with numerosity whereas array size correlated negatively with numerosity in half of the trials. In the other half, these relations were reversed (see Barth et al., 2006; Gilmore et al., 2010; Xenidou-Dervou et al., 2013; Xenidou-Dervou et al., 2014).

The *symbolic* version was identical to the nonsymbolic, only now the dot stimuli were replaced by the corresponding Arabic numerals (Gilmore et al., 2007; Xenidou-Dervou et al., 2013). Children were asked to respond to the question "Who got more stickers, Sarah or Peter?" by pressing the red or the blue response box in front of them (Fig. 1B).

2.3.2. General mathematics achievement

In the Netherlands, children's progress in primary school is monitored with the administration of the CITO tests. We acquired children's ability scores on the CITO Mathematics tests (in Dutch: CITO Rekenen-Wiskunde), which were assessed at the end of Grade 2 (June). The CITO math tests consist of many problems that cover a wide range of math domains: e.g., numbers and number relations, mental arithmetic (addition, subtraction, multiplication and division), complex applications (i.e. mostly more than one operation per problem), measurement (e.g. weight, length, time). This series of tests have been demonstrated to have good psychometric properties and high reliability (see Janssen, Verhelst, Engelen, & Scheltens, 2010).

2.3.3. Control measures

2.3.3.1. IQ. Children's non-verbal intelligence was assessed at the beginning of kindergarten with the Raven's Colored Progressive Matrices (Raven, Raven, & Court, 1998) in a group-testing session. This well-known test entails visual patterns with increasing difficulty. In each trial, a pattern is presented with a missing piece. The participant's task is to identify the missing piece, which will complete the design, out of six pieces. Children's raw scores on this test were used.

2.3.3.2. Working Memory. We used the Dutch version of six tasks adapted from the Automated WM assessment battery (AWMA; Alloway, 2007; Alloway, Gathercole, Willis, & Adams, 2004) that are often used to tap children's capacity on three subcomponents of WM, namely the Phonological Loop, the Visuospatial Sketchpad and the Central Executive component (Baddeley, 2012; Friso-van den Bos, Kroesbergen, & van Luit, 2014; Passolunghi & Lanfranchi, 2012; Xenidou-Dervou et al., 2013). We were interested in controlling for all aspects of WM capacity; therefore, we used both math-specific, i.e., entailing numbers, and not math-specific WM tasks. Moreover, we assessed both the ability of only retaining visuospatial or phonological information (VSSP and PL, respectively), as well as their interaction with the CE component (Repovs & Baddeley, 2006). Children's WM capacities were assessed in kindergarten, grade 1 and grade 2.

2.3.3.3. Visuospatial Sketchpad (VSSP). The VSSP component of WM was assessed with the "Cross Matrix". The *Cross Matrix* is identical to the well-known Dot Matrix of the AWMA battery; only in this version dots were replaced with crosses in order to exclude confounding factors with our nonsymbolic tasks that entailed dots. In this task the child saw a 4×4 matrix in which a cross appeared and disappeared. The child was instructed to remember the location of the cross and point to the correct box where the cross had previously appeared. A point was awarded for every correct response. After four correctly responded trials, the child was automatically advanced to the next level of difficulty, where one extra cross appeared. The task's levels of difficulty ranged from one up to five series. A correct response necessitated recalling correctly both the location and the right order in which the crosses appeared on the screen. If the child made three errors within one level of difficulty, the task was automatically terminated. The outcome measure entailed the total number of correct responses.

2.3.3.4. Phonological loop (PL). Children's PL capacity was assessed with the: "Word Recall Forward" and the "Digit Recall Forward" task. In the *Word Recall*, the child heard a series of unrelated, high frequency words, which had to be later recalled correctly and in the right order. The *Digit Recall* task was the same as the word recall task, only now the child had to recall correctly and in the right order digits instead of words. Scoring and task progression rules were identical to the VSSP tasks.

2.3.3.5. Central executive (CE). The CE can be fractionated on the basis of the information that is being manipulated within ones WM (Repovs & Baddeley, 2006). We, therefore, used three tasks to assess the children's CE capacity: the "Word Recall Backwards" (for not math-specific phonological information) and the "Digit Recall Backwards" task (for math-specific phonological information) and the Odd One Out (for visuospatial information). The *Word Recall Backwards* and *Digit Recall Backwards* tasks were similar to the Word Recall Forward and Digit Recall Forward tasks; only now the child was required to recall the words in the reversed order. The *Odd One Out* task started with the child seeing three shapes and was asked to point to the shape that differed from the other two. The shapes would then disappear from the screen and the child had to point to the location of the previously located odd one out shape. With increasing levels of difficulty, the set of presented shapes increased. A response was registered as correct when the child pointed out correctly and in the correct order the location of the odd shapes. Task progression rules were identical to the Cross Matrix task.

3. Results

Table 2 depicts descriptive statistics on math achievement performance and the control measures. Table A1 (Appendix) depicts the correlations between accuracy and RT in the four magnitude comparison measures: nonsymbolic and symbolic sequential-large and simultaneous-small across the three years of measurement (kindergarten, grade 1 and grade 2), and math achievement. There was no indication for an accuracy-RT trade-off between these four measures and general math achievement. Also, the correlations between the accuracy and RT data in the sequential-large measures did not indicate any accuracy-RT trade-off. There was a small indication of such a trade-off amongst the data of the simultaneous-small measures. Most correlations in this case were relatively small (the highest one was $r = 0.22$). In large samples such as the current one, however, even small correlations become significant. Taken together, these results suggested that accuracy and RT data should be examined separately in the subsequent analyses. In Table A1, one notices moderate to large correlations (Cohen, 1992) amongst corresponding longitudinal nonsymbolic and symbolic measures in both accuracy and RT.

3.1. Comparing developmental trajectories

To compare the developmental trajectories of nonsymbolic and symbolic simultaneous-small and sequential-large magnitude comparison *accuracy*, we conducted a 3 (Year: kindergarten, grade 1, grade 2) x 2 (Measure: simultaneous-small and sequential-large) x 2 (Stimulus: nonsymbolic and symbolic) repeated measures ANOVA. In the case of violation of the assumption of sphericity, degrees of freedom were corrected using Greenhouse-Geisser estimates. As expected, we found a significant Year by Task by

Table 2
Means (and SDs) of the control measures and the dependent variable (mathematics achievement).

Measure	Kindergarten		Grade 1		Grade 2		
	M (SD)	Max	M (SD)	Max	M (SD)	Max	Th.Max
IQ	21.24 (5.08)	34				34	36
Dot Matrix	9.97 (2.73)	16	13.24 (3.00)	20	15.29 (3.03)	24	24
Odd One Out	8.47 (2.88)	15	11.39 (2.56)	20	13.27 (2.58)	20	24
Word RF	13.89 (2.48)	22	14.98 (2.65)	28	15.77 (2.48)	21	28
Word RB	4.96 (1.82)	13	6.14 (2.11)	18	6.81 (2.26)	14	24
Digit RF	14.02 (2.42)	21	15.83 (2.41)	24	17.04 (2.48)	25	32
Digit RB	4.54 (1.64)	12	6.20 (2.03)	14	7.11 (2.32)	15	28
Maths Achievement					67.06 (14.73)	109	109

Note. RF = Recall Forward, RB = Recall Backwards, Th. Max = theoretical maximum score.

Stimulus interaction effect, F (1.83, 598.48) = 111.05, p < 0.001, $\eta_p^2 = 0.25$ (Fig. 2A and B). Therefore, the two measures and the two stimuli formats demonstrated different developmental trajectories. To unravel the simple effects, two additional analyses were conducted for each measure (simultaneous-small and sequential-large). For the simultaneous-small tasks, results demonstrated only main effects of Year, F (1.54, 509.6) = 133.38, p < 0.001, $\eta_p^2 = 0.29$ and Stimulus, F (1, 330) = 141.28, p < 0.001, $\eta_p^2 = 0.30$ (Fig. 2A), indicating that performance in the nonsymbolic and the symbolic task increased in a similar manner across the grades. Inspecting Fig. 2A, though, one notices that there was a ceiling effect in this measure. For the sequential-large tasks, results showed significant main effects for Year, F (1.96, 643.86) = 447.95, p < 0.001, $\eta_p^2 = 0.58$, and Stimulus, F (1, 328) = 135.61, p < 0.001, $\eta_p^2 = 0.29$, but also an interaction effect, F (1.92, 631.17) = 135.61, p < 0.001, $\eta_p^2 = 0.35$ (Fig. 2B). Thus, as expected, nonsymbolic and symbolic performance in the large-sequential task format demonstrated different developmental trajectories. Performance in the symbolic condition underwent larger developmental growth than its nonsymbolic counterpart.

The same analyses were conducted with the four measures' *RT* data. Once again, the 3 × 2 x 2 repeated measures ANOVA showed a significant 3-way interaction: Year by Task by Stimulus, F (1.81, 594.59) = 7.86, p = 0.001, $\eta_p^2 = 0.02$. For the simultaneous-small task, RT results demonstrated significant Year, F (1.59, 526.04) = 309.05, p < 0.001, $\eta_p^2 = 0.48$, and Stimulus, F (1, 330) = 164.95, p < 0.001, $\eta_p^2 = 0.33$, main effects but this time also the expected Year by Stimulus interaction effect, F (1.63, 539.34) = 81.22, p < 0.001, $\eta_p^2 = 0.20$ (Fig. 2C). For the sequential-large, as in the case of the accuracy data, we found significant Year, F (1.78, 583.18) = 270.21, p < 0.001, $\eta_p^2 = 0.45$, and Stimulus, F (1, 328) = 156.1, p < 0.001, $\eta_p^2 = 0.32$, as well as the expected Year by

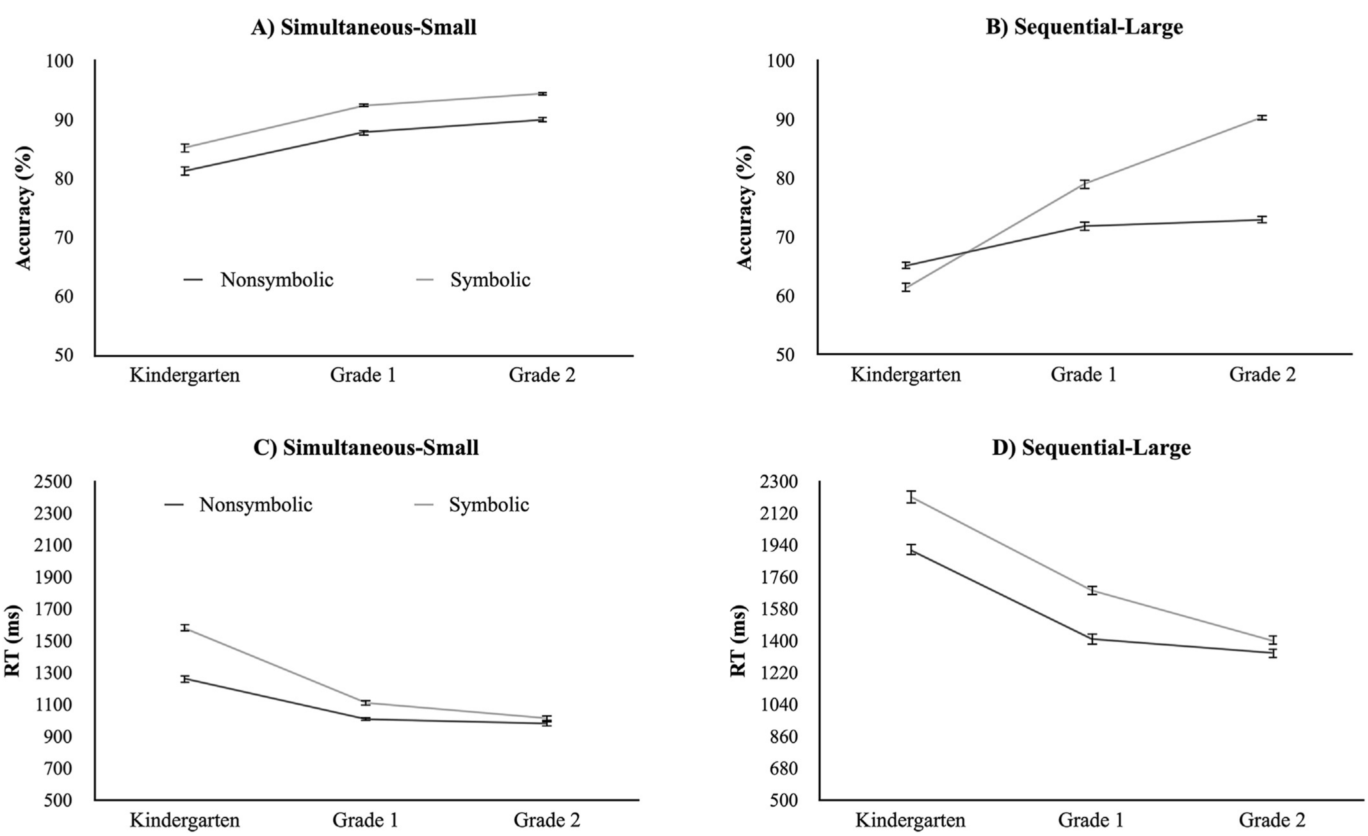

Fig. 2. Development in nonsymbolic and symbolic simultaneous-small and sequential-large comparison accuracy (A, B) and RT respectively (C, D) from kindergarten up to grade 2. Nonsymbolic and symbolic magnitude processing demonstrated different developmental trajectories. Also, the two measures (simultaneous-small vs. Sequential-large) had different developmental trajectories.

Stimulus interaction effect, F (1.83, 600.75) = 17.92, p < 0.001, $\eta_p^2 = 0.05$ (Fig. 2D). Thus, in line with our hypotheses, our findings confirmed that the two measures demonstrate different developmental trajectories. Furthermore, as hypothesized, nonsymbolic and symbolic magnitude comparison processing in both measures demonstrated different developmental trajectories.

3.2. Predicting future mathematics achievement

For the research question 2A, we sought to identify how the unique predictive power of the different nonsymbolic and symbolic magnitude processing skills changes across grades when predicting distant maths achievement. Therefore, we conducted a series of multiple linear regression analyses, one for each year, controlling for age, initial IQ and performance in the PL, VSSP and CE WM tasks in the respective year. One regression analysis for each grade was conducted by entering all variables in one step. This was done separately for magnitude comparison accuracy (Table 3) and RT data (Table 4). In the case of the accuracy scores, F-tests indicated that all models significantly explained variance in grade 2 general math achievement: kindergarten, $F(12, 298) = 15.12$, $p < 0.001$, Adj. $R^2 = 0.35$, grade 1, $F(12, 299) = 17.16$, $p < 0.001$, Adj. $R^2 = 0.38$ and grade 2, F (12, 293) = 13.46, p < 0.001, Adj. $R^2 = 0.33$. In kindergarten and grade 1, both nonsymbolic and symbolic sequential-large magnitude processing uniquely predicted distant math achievement above and beyond WM skills and IQ (Table 3). We further compared the nonsymbolic and symbolic regression coefficients (Neter, Wasserman, & Kutner, 1985) and found that symbolic sequential-large was a better predictor of math achievement both in kindergarten (p = 0.000) as well as in grade 1 (p = 0.011) than the nonsymbolic one. In grade 2, out of the four magnitude comparison predictors assessed in the beginning of grade 2, only symbolic sequential-large performance explained unique variance in math achievement at the end of the grade.

Also in the case of the RTs, F-tests showed that all models significantly explained variance in grade 2 general math achievement: kindergarten, $F(12, 298) = 10.89$, $p < 0.001$, Adj. $R^2 = 0.28$, grade 1, $F(12, 299) = 12.95$, $p < 0.001$, Adj. $R^2 = 0.34$ and grade 2, F (12, 293) = 15.69, p < 0.001, Adj. $R^2 = 0.37$. In this case, only the symbolic magnitude comparison predictors reached significance (Table 4). Speed in comparing small number digits in kindergarten, and large numbers in grade 1 and grade 2 uniquely predicted distal math achievement above and beyond nonsymbolic magnitude processing speed, WM resources and initial IQ.

Inspecting Table 3 one notices that not only does symbolic comparison, as assessed with the sequential-large measure, appears to be an important predictor of children's future mathematics in every grade, but also its regression coefficient seems to be comparable to that of initial IQ. So, we further compared the regression coefficients of IQ and symbolic sequential-large

Table 3
Accuracy (%) Per Year on the Magnitude Comparison Measures Predicting Maths Achievement at the End of Grade 2, While Controlling for WM capacities on the Given Year, initial IQ and Age.

Predictors	Kindergarten		Grade 1		Grade 2	
	B (SE)	β	B (SE)	β	B (SE)	β
Age	−6.32 (2.03)	−0.15**	−1.98 (2.06)	−0.04	1.30 (2.19)	0.03
IQ	0.57 (0.16)	0.19***	0.51 (0.16)	0.17***	0.76 (0.16)	0.25***
Word RF	−0.22 (0.37)	−0.04	0.38 (0.32)	0.07	−0.04 (0.38)	−0.01
Word RB	1.03 (0.45)	0.12*	0.31 (0.36)	0.04	0.76 (0.37)	0.12*
Digit RF	1.23 (0.39)	0.19**	0.38 (0.37)	0.06	0.85 (0.40)	0.14*
Digit RB	1.15 (0.46)	0.13*	0.33 (0.38)	0.04	0.68 (0.38)	0.11[a]
Dot Matrix	0.29 (0.28)	0.05	0.70 (0.26)	0.14**	0.34 (0.28)	0.07
Odd One Out	0.14 (0.25)	0.03	0.46 (0.30)	0.08	0.62 (0.32)	0.11[a]
Nonsymbolic SimS	0.10 (0.06)	0.10[a]	0.07 (0.08)	0.04	0.04 (0.10)	0.02
Symbolic SimS	0.02 (0.06)	0.02	0.07 (0.10)	0.03	0.09 (0.15)	0.03
Nonsymbolic SeqL	0.13 (0.06)	**0.12***	0.13 (0.06)	**0.10***	0.09 (0.06)	0.08
Symbolic SeqL	0.23 (0.05)	**0.24*****	0.34 (0.05)	**0.33*****	0.31 (0.09)	**0.18*****

Note. General math achievement at the end of grade 2 was the dependent variable for all models. Significant magnitude comparison predictors in bold. SeqL = Sequential-Large, SimS = Simultaneous-Small, RF = Recall Forward, RB = Recall Backwards, ***p ≤ 0.001, **p ≤ 0.01, *p ≤ 0.05.
[a] = p < 0.10.

Table 4
Average RT (ms) Per Year on the Magnitude Comparison Measures Predicting Maths Achievement at the End of Grade 2, While Controlling for WM capacities on the Given Year, initial IQ and Age.

Predictors	Kindergarten		Grade 1		Grade 2	
	B (SE)	β	B (SE)	β	B (SE)	β
Age	−4.08 (2.14)	−0.10[a]	−1.73 (2.17)	−0.04	1.18 (2.14)	0.03
IQ	0.72 (0.16)	0.24***	0.64 (0.16)	0.21***	0.71 (0.15)	0.23***
Word RF	−0.08 (0.40)	−0.01	0.33 (0.34)	0.06	0.06 (0.36)	0.01
Word RB	1.13 (0.48)	0.14*	0.32 (0.38)	0.05	0.75 (0.37)	0.11*
Digit RF	1.26 (0.42)	0.20**	0.78 (0.38)	0.12*	0.85 (0.39)	0.14*
Digit RB	1.30 (0.49)	0.14**	0.74 (0.40)	0.10+	0.82 (0.37)	0.13*
Dot Matrix	0.44 (0.30)	0.08	0.74 (0.28)	0.15**	0.26 (0.28)	0.05
Odd One Out	0.11 (0.26)	0.02	0.61 (0.32)	0.11[a]	0.67 (0.31)	0.12*
Nonsymbolic SimS	0.00 (0.00)	0.07	−0.01 (0.00)	−0.10[a]	0.00 (0.00)	0.05
Symbolic SimS	−0.01 (0.00)	**−0.16****	−0.00 (0.00)	−0.06	−0.01 (0.00)	-0.12[a]
Nonsymbolic SeqL	−0.00 (0.00)	−0.05	−0.00 (0.00)	−0.06	0.00 (0.00)	0.01
Symbolic SeqL	0.00 (0.00)	−0.01	−0.00 (0.00)	**−0.13***	−0.01 (0.00)	**−0.25*****

Note. General math achievement at the end of grade 2 was the dependent variable for all models. Significant magnitude comparison predictors in bold. SeqL = Sequential-Large, SimS = Simultaneous-Small, RF = Recall Forward, RB = Recall Backwards, ***p ≤ 0.001, **p ≤ 0.01, *p ≤ 0.05.
[a] = p < .10.

comparison accuracy in each grade. In kindergarten, symbolic comparison was a significantly stronger predictor than that of IQ ($p = 0.042$). In grade 1, the two predictors were equally important ($p = 0.312$), and in grade 2 IQ was a stronger predictor than symbolic comparison ($p = 0.010$).

The previous regression analyses examined how magnitude comparison measures uniquely contributed to future mathematical achievement at each grade (kindergarten, grade 1 and grade 2). However, given the developmental design of our study, we were also interested in examining, whether magnitude comparison performance in the subsequent years (grade 1 and 2) improved the prediction of future maths achievement over and above the kindergarten predictors (Research question 2B). Also, which of the magnitude comparison measures uniquely contributed to children's future mathematical achievement across all grades, taking into account the fact that grade 1 magnitude comparison correlates with kindergarten magnitude comparison etc. (Research question 2C). In other words, we aimed to control for the shared variance that each nonsymbolic or symbolic predictor had with the same predictor in the other years. Therefore, we conducted two hierarchical linear regression analyses, one for magnitude comparison accuracy, and one for magnitude comparison RT, controlling for age, initial IQ and children's developing WM abilities (Table 5). In these models, in Step 1 we entered age, IQ and children's WM capacities and in each next step we entered their performance in the four magnitude comparison measures in each year (Kindergarten: Step 2, Grade 1: Step 3, Grade 2: Step 4). As we had multiple WM variables measured in each year, and we were only interested in controlling for WM as a general construct, we computed composite scores of performance across the three years in each WM task, i.e., one composite score for performance in the Dot Matrix task across kindergarten, grade 1 and grade 2, and similarly for the Odd One Out, the Digit Recall Forward, the Digit Recall Backwards, Word Recall Forward and Word Recall Backwards task.

Table 5
Results on the Last Step of the Hierarchical Regressions Focusing on How Accuracy (%) and RT Across All years on the Magnitude Comparison Measures Predict Future Maths Achievement, while Controlling for Developing WM Capacities, Initial IQ and Age.

Predictors	Magn. Accuracy		Magn. RT	
	B (Std. Error)	β	B (Std. Error)	β
Age	−4.26 (1.93)	−0.10*	−2.03 (2.00)	−0.05
IQ	0.43 (0.15)	0.14**	0.50 (0.16)	0.16**
Dot Matrix comp	0.83 (0.43)	0.11[a]	0.85 (0.45)	0.12[a]
Odd One Out comp	0.55 (0.45)	0.07	0.76 (0.46)	0.09
Word RF comp	−0.22 (0.52)	−0.03	−0.08 (0.54)	−0.01
Word RB comp	0.90 (0.56)	0.09	0.88 (0.60)	0.09
Digit RF comp	0.86 (0.54)	0.12	0.91 (0.56)	0.13
Digit RB comp	0.71 (0.58)	0.07	1.76 (0.61)	0.18**
Nonsymbolic SimS K	0.04 (0.06)	0.04	0.00 (0.00)	0.10[a]
Symbolic SimS K	−0.06 (0.06)	−0.06	−0.00 (0.00)	−0.09
Nonsymbolic SeqL K	0.12 (0.06)	**0.10***	0.00 (0.00)	−0.01
Symbolic SeqL K	0.13 (0.05)	**0.14****	0.00 (0.00)	0.00
Nonsymbolic SimS G1	−0.00 (0.08)	−0.00	−0.01 (0.00)	−0.10
Symbolic SimS G1	0.07 (0.10)	0.03	0.00 (0.00)	0.07
Nonsymbolic SeqL G1	0.03 (0.06)	0.02	−0.00 (0.00)	−0.03
Symbolic SeqL G1	0.26 (0.06)	**0.25*****	−0.00 (0.00)	−0.06
Nonsymbolic SimS G2	−0.04 (0.10)	−0.02	0.00 (0.00)	0.02
Symbolic SeqL G2	0.03 (0.14)	0.01	−0.01 (0.00)	-0.09
Nonsymbolic SeqL G2	0.03 (0.06)	0.03	0.00 (0.00)	0.04
Symbolic SeqL G2	0.20 (0.09)	**0.11***	−0.01 (0.00)	**−0.20****

Note. General mathematics achievement at the end of grade 2 was the dependent variable for both models. Significant magnitude comparison coefficients in bold. Comp = composite score, RF = Recall Forward, RB = Recall Backwards SeqL = Sequential-Large, SimS = Simultaneous-Small, K = kindergarten, G1 = grade 1, G2 = grade 2. ***p ≤ 0.001, **p ≤ 0.01, *p ≤ 0.05.
[a] = p < 0.10.

In the case of the model with the magnitude comparison accuracy data, results showed that the model significantly changed with each step [Step 1: $\Delta F\,(8, 289) = 20.24$, $p < 0.001$, Step 2: $\Delta F\,(4, 285) = 8.28$, $p < 0.001$, $\Delta F\,(4, 281) = 7.09$, $p < 0.001$], except for Step 4, $\Delta F\,(4, 277) = 1.50$, $p = 0.203$. Thus, magnitude comparison accuracy in Grade 2 did not improve the prediction of future maths achievement. Table 5 depicts the regression coefficients in the final step (Step 4) showing the unique contribution of each predictor across all years. All models explained variance in future mathematics achievement: Step 1: $F\,(8, 289) = 20.24$, $p < 0.001$, Adj. $R^2 = 0.34$, Step 2: $F\,(12, 285) = 17.61$, $p < 0.001$, Adj. $R^2 = 0.40$, Step 3: $F\,(16, 281) = 16.11$, $p < 0.001$, Adj. $R^2 = 0.45$, Step 4: $F\,(20, 277) = 13.28$, $p < 0.001$, Adj. $R^2 = 0.45$. Once again, it was evident (Table 5) that performance in the symbolic sequential-large measure in every grade (kindergarten, grade 1 and grade 2) uniquely predicted future mathematical achievement above and beyond domain-general capacities and nonsymbolic processing. Its regression coefficient appeared to be comparable to IQ, therefore we compared the coefficients of IQ with each symbolic sequential-large predictor and found that in all cases they were equally important: kindergarten ($p = 0.060$), grade 1 ($p = 0.298$) and grade 2 ($p = 0.151$). This time, only kindergarten nonsymbolic sequential-large comparison uniquely predicted future mathematics achievement.

With the magnitude comparison RT data, the model significantly changed with each step [Step 1: $\Delta F\,(8, 289) = 20.24$, $p < 0.001$, Step 2: $\Delta F\,(4, 285) = 2.35$, $p = 0.055$ (marginally significant), $\Delta F\,(4, 281) = 3.70$, $p = 0.006$, Step 4: $\Delta F\,(4, 277) = 3.54$, $p = 0.008$]. So, the addition of the RT magnitude predictors of both subsequent years significantly improved the prediction model. All models explained variance in future mathematics achievement: Step 1: $F\,(8, 289) = 20.24$, $p < 0.001$, Adj. $R^2 = 0.34$, Step 2: $F\,(12, 285) = 14.53$, $p < 0.001$, Adj. $R^2 = 0.35$, Step 3: $F\,(16, 281) = 12.23$, $p < 0.001$, Adj. $R^2 = 0.38$, Step 4: $F\,(20, 277) = 10.85$, $p < 0.001$, Adj. $R^2 = 0.40$. The two far right columns in Table 5 depict the regression coefficients in the final step of the regression (Step 4), showing the unique contribution of each predictor across all years. This time only RT in grade 2 symbolic sequential-large magnitude comparison uniquely predicted future mathematical achievement and its regression coefficient was significantly larger than that of IQ ($p = 0.002$).[1]

4. Discussion

The present study's findings shed further light onto the roles that nonsymbolic and symbolic magnitude comparison skills play in the transition from kindergarten to formal schooling (grade 1 and grade 2) and try to reconcile existing conflicting findings in the literature (for reviews see De Smedt et al., 2013; Feigenson et al., 2013). For the first time, a single large sample of children was assessed on two different commonly used nonsymbolic and symbolic magnitude comparison measures from kindergarten through to grade 2. We also assessed the children's IQ in kindergarten, their

[1] We also ran latent growth models with the intercept and slope in each magnitude comparison task (accuracy and RT) data as predictors of future mathematics achievement. The main outcomes were similar to the ones reported with the hierarchical regression analyses. In essence, latent growth modeling revealed that even though various nonsymbolic and symbolic magnitude measures correlated with children's individual differences at the kindergarten stage (i.e., initial status), when it came to potential for developmental change, only individual developmental growth in the symbolic sequential-large magnitude comparison measure correlated with children's future mathematical achievement. For clarity reasons, however, we only report the regression analyses results, where it was also possible to control for domain general capacities.

developing WM abilities (in kindergarten, grade 1 and grade 2), and their general mathematical achievement at the end of grade 2. Our results showed that: 1) As expected, with formal education, symbolic processing demonstrated larger developmental growth across the three grades than nonsymbolic processing. 2) Performance on the two different types of measures, widely used to assess nonsymbolic or symbolic magnitude comparison skills, also followed different developmental trajectories. This indicated that measures that differ on the basis of number ranges and design characteristics should not be addressed as interchangeable measures within the literature. Comparison measures, such as the simultaneous-small one in the present study, which include small numbers that are presented simultaneously, are easy for children and demonstrate a ceiling effect early on in development. 3) The predictive role of nonsymbolic and symbolic comparison skills dynamically changed across grades. Both nonsymbolic and symbolic magnitude comparison – as assessed with measures that include large numbers and sequential steps – uniquely predicted children's future mathematical achievement in kindergarten above and beyond IQ and WM abilities. With the start of formal mathematics education in school, however, symbolic comparison took over as the sole unique magnitude comparison predictor of future mathematics. In general, symbolic magnitude comparison was consistently a more robust and consistent predictor of future general mathematical achievement than nonsymbolic and its predictive power was mostly similar or even stronger to that of IQ.

4.1. Nonsymbolic versus symbolic developmental rates

The fact that we share a cognitive ability with other species – the “innate” ability to estimate abstract quantities in nature, i.e., the ANS – generates a lot of questions. How does this evolutionary ancient ontogenetic and phylogenetic cognitive system relate to our ability as humans to use symbols to represent quantities precisely? It has often been assumed that symbolic representations directly map one-to-one onto our readily accessible nonsymbolic representations (Lipton & Spelke, 2005; Mundy & Gilmore, 2009; Piazza & Izard, 2009). If that were the case then one may expect that symbolic and nonsymbolic magnitude processing would demonstrate, quantitatively similar developmental growth rates. Our results, however, with two different magnitude comparison measures appear to indicate that this may not be the case (see Lyons et al., 2012; Matejko & Ansari, 2016). Nonsymbolic and symbolic *magnitude comparison* processing demonstrated different developmental pathways as in the case of nonsymbolic and symbolic *approximate arithmetic* (Xenidou-Dervou et al., 2015a). Of course, non-parallel developmental curves do not automatically imply that the two abilities are completely disconnected. Perhaps the ANS partially influences symbolic magnitude at the early stages of development (Xenidou-Dervou et al., 2013), and this directionality may change later on in development (Noël & Rousselle, 2011) or both abilities could start to affect each other reciprocally (Mussolin, Nys, Leybaert, & Content, 2016). Nevertheless, our results appear to indicate that symbolic processing is influenced more by formal education and experience than nonsymbolic processing. Growth in symbolic magnitude comparison with education can be attributed to children's increasing understanding of the place-value system (Nuerk, Kaufmann, Zoppoth, & Willmes, 2004) and increase in mathematical knowledge in general, since the relationship between symbolic processing and mathematical achievement can be bidirectional (Case et al., 1997; Friso-van den Bos, Kroesbergen, et al., 2015). An alternative theoretical account for the observed larger developmental growth in symbolic comparison compared to nonsymbolic, could be that children do actually use the ANS for symbolic processing, but need to learn the Arabic numerals better to access it (Rousselle & Noël, 2007).

4.2. Different magnitude measures

Beyond the stimulus distinction though (nonsymbolic versus symbolic), another distinction between magnitude comparison skills is the type of measure used to assess the nonsymbolic or the symbolic system. Researchers so far have been using various different types of measures interchangeably. However, these measures differ on the basis of multiple design characteristics. In this study, we used two of such widely used types of measures, which differed both on the basis of numerical range (1–9 vs. 6 to 70) as well as the presentation format (i.e., numbers presented simultaneously or in sequential steps). We assumed that this distinction might be a source of the inconsistent findings evidenced in the literature. Indeed, our findings revealed that performance in the two different measures, both in the case of nonsymbolic processing as well as in symbolic processing, follow different developmental trajectories. Specifically, the measures that make use of numbers ranging from 1 to 9 presented in simultaneous steps, quickly reached ceiling effects for accuracy (see Fig. 2A). In contrast, larger developmental growth was evident in children's performance in the sequential-large measures (nonsymbolic and symbolic, with numbers ranging from 6 to 70 presented in sequential steps). In the literature, the simultaneous-small measures have often been used, and ceiling effects are thought to be circumvented with the use of RT data. However, our results showed that accuracy and RT data followed different developmental trajectories. The evidently disconnected developmental trajectories of the two different types of measures raise concerns as to how the nonsymbolic and the symbolic system are currently being assessed. The fact that small numbers are processed differently than large numbers should not be forgotten (Feigenson, Dehaene, & Spelke, 2004; Nuerk et al., 2004). Also, different number ranges employ different WM resources (see Xenidou-Dervou et al., 2015b, 2014) and the role of WM depends on the design characteristics of a given cognitive task and the age of the participants (Friso-van den Bos et al., 2013; Rasmussen & Bisanz, 2005; Simmons et al., 2012; Xenidou-Dervou et al., 2015b).

4.3. Magnitude comparison skills as longitudinal predictors of mathematics

As outlined in the introduction, due to the inconsistent findings across the literature the unique predictive roles of nonsymbolic and symbolic processing skills have been unclear (for reviews see De Smedt et al., 2013; Feigenson et al., 2013): Which ability explains children's individual differences in mathematical achievement at the early stages of development? Nonsymbolic, symbolic magnitude processing, or both? To address this question, we first ran regression analyses to identify the unique predictive role of the magnitude comparison skills for *each year* (kindergarten, grade 1 and grade 2). With the *accuracy* data (Table 3), we found that in kindergarten and grade 1, both nonsymbolic and symbolic sequential-large magnitude comparison played a unique role in predicting distant math achievement (Gilmore et al., 2010; Hornung et al., 2014). They were unique longitudinal predictors above and beyond all WM capacities, age and IQ. Symbolic sequential-large magnitude comparison was consistently a stronger predictor compared to its nonsymbolic counterpart and noticeably its unique predictive power was even stronger than IQ in kindergarten and similar to IQ in grade 1. In grade 2, symbolic sequential-large magnitude comparison was the only magnitude comparison skill that explained unique variance in children's general mathematical achievement longitudinally beyond domain-general capacities.

Secondly, we examined which magnitude comparison skill explained individual differences in children's future mathematical achievement *across all three years* (Table 5), controlling this way for test-retest effects. Although in the previous analyses we found several magnitude comparison accuracy variables predicting distant mathematics achievement year after year, the hierarchical regressions showed whether this meant that the predictive power of the model (in terms of explained variance) became stronger with the addition of each subsequent year. This was true for the addition of the magnitude comparison predictors of grade 1. This was probably the result of the increased importance of the symbolic comparison predictors. The inclusion of the predictors of grade 2, however, did not yield a gain in the predictive power of the model. Although the separate regression analysis of the last year showed the importance of the symbolic comparison predictors in that year, the inclusion of the magnitude comparison predictors of the last year apparently did not add new information to the prediction model. In other words, the predictive strength of the model stabilized in grade 1. This time nonsymbolic sequential-large performance explained unique variance only in kindergarten and its predictive power was relatively small (see also Schneider et al., 2016). Symbolic sequential-large magnitude comparison, however, took over, as expected, as a robust unique predictor of future mathematics across all three years (De Smedt et al., 2013) and its predictive strength was similar to that of IQ. Although domain-general capacities were only used as control measures in the present study, it should be noted that of course, as expected, IQ and different WM abilities – especially the CE component – were consistently significant predictors of children's future mathematical achievement (De Smedt, Janssen, et al., 2009; DeStefano & LeFevre, 2004; Geary et al., 2009; Geary, Hoard, & Nugent, 2012; Raghubar et al., 2010).

Our regression findings also empirically verified the assumption that RT and accuracy data yield different patterns of results (De Smedt et al., 2013). With respect to the speed of comparing nonsymbolic or symbolic magnitudes (*RT*), only the symbolic measures explained individual differences in mathematical achievement longitudinally. When regressions were run for *each year* (Table 4), the simultaneous-small symbolic measure appeared to play a unique role in kindergarten and the symbolic sequential-large in grades 1 and 2, indicating a developmental shift from small to large numbers with the start of formal schooling. In the hierarchical regression analysis, where data from *all three years* were entered step by step, the prediction model improved with the addition of each subsequent year (Table 5). In this case, only grade 2 RT in the symbolic-sequential large magnitude comparison measure uniquely explained children's future mathematical achievement. Its predictive power was even stronger than that of IQ.

The fact that kindergarten nonsymbolic performance predicted future mathematics achievement above and beyond all the assessed domain-general capacities and symbolic processing, appears to support the assumption that the ANS may play a unique predictive role in mathematics achievement at the initial stages of development. Alternatively, research suggests that nonsymbolic effects may in fact be an artefact of the inhibitory demands entailed in nonsymbolic comparison tasks (Gilmore, et al., 2013). Although we assessed and controlled for children's central executive WM capacity, which as a cognitive construct is considered to incorporate inhibition abilities (Pureza, Jacobsen, Oliveira, & Fonseca, 2011), we did not assess children's inhibition skills per se. Future research should address this limitation and examine whether nonsymbolic sequential-large performance can uniquely predict children's mathematics achievement beyond any inhibitory control capacities.

Nevertheless, our findings clearly support the predominance of symbolic processing and its growth as a unique, robust and consistent predictor of children's future mathematics achievement (De Smedt et al., 2013; Lyons et al., 2014). This finding supports the assumption that good knowledge of the numerical meaning of symbolic numbers, rather than their nonsymbolic representations, is a fundamental precursor of children's mathematical development (De Smedt et al., 2009; Sasanguie et al., 2011). Alternatively, perhaps the role of symbolic processing could be attributed to children's individual differences in their ability to access the ANS from symbols and not their ability to distinguish numerosities per se (Rouselle & Noël, 2007). It should be noted, however, that this assumption was not tested in this study and it does not fully explain our results since we found nonsymbolic processing to be a significant unique predictor too at the kindergarten stage, beyond their symbolic skills (although see the aforementioned alternative explanation concerning the role of inhibition). To be able to address directly the question of whether children differ on the basis of their ability to access the ANS via symbolic processing, future research should examine longitudinally children's ability to map number symbols to nonsymbolic quantities and the other way around using mapping tasks (Mundy & Gilmore, 2009).

Interestingly, the regression results further revealed how previous incompatible findings could be attributed to the type of magnitude measure that is used to assess nonsymbolic or the symbolic abilities. As in previous studies, we found no evidence for the nonsymbolic *simultaneous-small* measure uniquely predicting mathematics achievement at any stage with either accuracy or RT data (Bartelet et al., 2014; Holloway & Ansari, 2009; Lyons et al., 2014; Sasanguie et al., 2011, 2014, 2013). Only the *sequential-large* nonsymbolic measure played a unique predictive role in kindergarten. In the case of the symbolic measures, the simultaneous-small one was once again a poor predictor compared to its sequential-large counterpart. As mentioned earlier, performance in both simultaneous-small measures demonstrated ceiling effects early on in development (see Fig. 2A) and thus could not explain individual differences in children's mathematical achievement. But RT in these measures did not prove to be better predictors either, with the exception only of RT in the symbolic simultaneous-small measure in kindergarten. It should be reiterated that the focus of this study was not the sequential vs. simultaneous, or small vs. large distinction; future research should experimentally address the cognitive mechanisms underlying performance in these different types of measures. Nevertheless, our findings indicate that performance in the nonsymbolic and symbolic sequential-large measures are better predictors of children's future mathematical achievement than their simultaneous-small counterparts.

4.4. Concluding remarks

The “nonsymbolic versus symbolic” debate is actually similar to the “nature versus nurture” debate. Admittedly, the assumption that we have an “innate” ability, the ANS, to estimate and manipulate nonsymbolic quantities in nature and that this ability may foster our mathematical achievement comprises a very compelling story (Dehaene, 2011; Feigenson et al., 2013, 2004; Starr et al., 2013). On the other hand, the assumption that nonsymbolic processing does not play a predictive role in the early developmental steps of mathematical achievement and what is of primary importance is how well children learn to compare symbolic numerals (De Smedt et al., 2009, 2013; Lyons et al., 2014; Noël & Rousselle, 2011; Sasanguie et al., 2014), is also quite a compelling theoretical account primarily because symbolic skills can potentially be easier to enhance by learning and instruction than innate skills. The present study demonstrated that the inconsistent findings in the literature could be attributed to the developmental stage one is examining and the type of measure or data (accuracy or RT) one uses. The ANS, as an intuitive, readily accessible, nonverbal cognitive system, may play a moderate unique role in the development of general mathematics achievement only until the start of formal schooling. At this early developmental stage, it appears to be a unique predictor of children's future mathematical achievement beyond IQ and WM. Symbolic processing, however, appears to have an independent developmental growth rate, which is affected more by development and education than nonsymbolic processing and it is a robust and consistent precursor of children's future mathematics achievement across all three grades (kindergarten, grade 1 and grade 2) above and beyond domain-general capacities such as WM abilities and IQ. Contrary to the ANS, symbolic processing and growth necessitates the coordination of the multiple meanings of number (Case et al., 1997; Griffin, Case, & Siegler, 1994), i.e., knowledge of Arabic digits, their order, their phonological representations and place-value knowledge (Göbel, Watson, Lervåg, & Hulme, 2014; Lyons & Beilock, 2011; Sasanguie & Reynvoet, 2014; Xenidou-Dervou et al., 2015a). Notably, we found that its predictive power was comparable and even stronger at times to that of kindergarten IQ.

These findings have important implications for educational assessment and practice. Our results suggest that, besides domain general capacities such as IQ, symbolic magnitude comparison skills could potentially be used as a screening tool for identifying children with difficulties in mathematics. Also, future research should further examine with experimental studies whether improvement in mathematical achievement is achieved after training children's symbolic magnitude comparison skills (for a review see De Smedt et al., 2013). The fact that symbolic processing skills are likely affected by development and education and predict children's future mathematical achievement across all three grades (kindergarten up to grade 2), implies that the enhancement of this skill could potentially influence their future general mathematical achievement.

Acknowledgements

The authors would like to thank the entire interlinked Math-Child team and all the participating children, parents, teachers and schools. This research was conducted as part of a broader NWO (National Dutch Organization for Scientific Research) funded project; Grant Number: PROO 411 07 111.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.learninstruc.2016.11.001.

References

Alloway, T. P. (2007). *Automated working memory assessment*. London, UK: Pearson Assessment (Translated and reproduced by permission of Pearson Assessment).

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology, 87*, 85–106. http://dx.doi.org/10.1016/j.jecp.2003.10.002.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829–839. http://dx.doi.org/10.1038/nrn1201.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29. http://dx.doi.org/10.1146/annurev-psych-120710-100422.

Bartelet, D., Vaessen, A., Blomert, L., & Ansari, D. (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *Journal of Experimental Child Psychology, 117C*, 12–28. http://dx.doi.org/10.1016/j.jecp.2013.08.010.

Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005). *Abstract number and arithmetic in preschool children*. http://dx.doi.org/10.1073/pnas.0505512102. Proceedings of the National Academy of Sciences of the United States of America, 102, 14116–21.

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition, 98*, 199–222. http://dx.doi.org/10.1016/j.cognition.2004.09.011.

Case, R., Okamoto, Y., Griffin, S., McKeough, A., Bleiker, C., Henderson, B., et al. (1997). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development, 61*, 1–295. http://dx.doi.org/10.2307/1166077.

Cohen, J. (1992). A power primer. *Psychological bulletin, 112*(1), 155.

Collom, R., Escorial, S., Shih, P. C., & Privado, J. (2007). Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents. *Personality and Individual Differences, 42*, 1503–1514. http://dx.doi.org/10.1016/j.paid.2006.10.023.

De Smedt, B., & Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology, 108*, 278–292. http://dx.doi.org/10.1016/j.jecp.2010.09.003.

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology, 103*, 186–201. http://dx.doi.org/10.1016/j.jecp.2009.01.004.

De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, 1–8. http://dx.doi.org/10.1016/j.tine.2013.06.001.

De Smedt, B., Verschaffel, L., & Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *Journal of Experimental Child Psychology, 103*, 469–479. http://dx.doi.org/10.1016/j.jecp.2009.01.010.

Dehaene, S. (2011). *The number Sense: How the mind creates mathematics, revised and updated edition*. USA: Oxford University Press. Retrieved from http://www.amazon.com/The-Number-Sense-Creates-Mathematics/dp/0199753873.

Dehaene, S., Izard, V., & Piazza, M. (2005). *Control over non-numerical parameters in numerosity experiments*. Unpublished manuscript (available on www.unicog.org).

DeStefano, D., & LeFevre, J. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology, 16*, 353–386. http://dx.doi.org/10.1080/09541440244000328.

Dougherty, C. (2003). Numeracy, literacy and earnings: Evidence from the national longitudinal survey of youth. *Economics of Education Review, 22*, 511–521. http://dx.doi.org/10.1016/S0272-7757(03)00040-2.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*, 307–314. http://dx.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*, 74–79. http://dx.doi.org/10.1111/cdep.12019.

Friso-van den Bos, I., Kroesbergen, E. H., & van Luit, J. E. H. (2014). Number sense in kindergarten children: Factor structure and working memory predictors. *Learning and Individual Differences, 33*, 23–29. http://dx.doi.org/10.1016/j.lindif.2014.05.003.

Friso-van den Bos, I., Kroesbergen, E. H., Van Luit, J. E. H., Xenidou-Dervou, I., Jonkman, L. M., Van der Schoot, M., et al. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of Experimental Child Psychology, 134*, 12–29. http://dx.doi.org/10.1016/j.jecp.2015.02.002.

Friso-van den Bos, I., van der Ven, S. H. G., Kroesbergen, E. H., & van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review, 10*, 29–44. http://dx.doi.org/10.1016/j.edurev.2013.05.003.

Friso-van den Bos, I., Van Luit, J. E. H., Kroesbergen, E. H., Xenidou-Dervou, I., van Lieshout, E. C. D. M., van der Schoot, M., & Jonkman, L. M. (2015). Pathways of Number Line Development in Children. *Zeitschrift Für Psychologie, 223*(2), 120–128. http://doi.org/10.1027/2151-2604/a000210.

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

Geary, D. C., Hoard, M. K., & Nugent, L. (2012). Independent contributions of the central executive, intelligence, and in-class attentive behavior to developmental change in the strategies used to solve addition problems. *Journal of Experimental Child Psychology, 113*, 49–65. http://dx.doi.org/10.1016/j.jecp.2012.03.003.

Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology. General, 141*, 642–648. http://dx.doi.org/10.1037/a0026218.

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., et al. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS One, 8*(6), e67374. http://dx.doi.org/10.1371/journal.pone.0067374.

Gilmore, C., Attridge, N., De Smedt, B., & Inglis, M. (2014). Measuring the approximate number system in children: Exploring the relationships among different tasks. *Learning and Individual Differences, 29*, 50–58. http://dx.doi.org/10.1016/j.lindif.2013.10.004.

Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *The Quarterly Journal of Experimental Psychology, 64*, 2099–2109.

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature, 447*(7144), 589–591. http://dx.doi.org/10.1038/nature05850.

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*, 394–406. http://dx.doi.org/10.1016/j.cognition.2010.02.002.

Göbel, S. M., Watson, S. E., Lervåg, A., & Hulme, C. (2014). Children's arithmetic development: It is number knowledge, not the approximate number sense, that counts. *Psychological Science, 25*, 789–798. http://dx.doi.org/10.1177/0956797613516471.

Griffin, S. A., Case, R., & Siegler, R. S. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure. In K. McGilly (Ed.), *Classroom Lessons: Integrating cognitive theory and classroom practice* (pp. 25–49). , Cambridge, MA: MIT Press.

Gullick, M. M., Sprute, L. a., & Temple, E. (2011). Individual differences in working memory, nonverbal IQ, and mathematics achievement and brain mechanisms associated with symbolic and nonsymbolic number processing. *Learning and Individual Differences, 21*, 644–654. http://dx.doi.org/10.1016/j.lindif.2010.10.003.

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*, 1457–1465. http://dx.doi.org/10.1037/a0012682.

Holloway, I. D., & Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. *Developmental Science, 11*, 644–649. http://dx.doi.org/10.1111/j.1467-7687.2008.00712.x.

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology, 103*, 17–29. http://dx.doi.org/10.1016/j.jecp.2008.04.001.

Hornung, C., Schiltz, C., Brunner, M., & Martin, R. (2014). Predicting first-grade mathematics achievement: The contributions of domain-general cognitive abilities, nonverbal number sense, and early number competence. *Frontiers in Psychology, 5*, 272. http://dx.doi.org/10.3389/fpsyg.2014.00272.

Inglis, M., Attridge, N., Batchelor, S., & Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: But only in children. *Psychonomic Bulletin & Review, 18*, 1222–1229. http://dx.doi.org/10.3758/s13423-011-0154-1.

Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). Wetenschappelijke verantwoordning van de toetsen LOVS Rekenen-Wiskunde voor groep 3 tot en met 8 [Scientific justification of the mathematics test for Grade 1 until Grade 6]. Arnhem, Netherlands: Cito.

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*, 1292–1300. http://dx.doi.org/10.1111/j.1467-7687.2011.01080.x.

Lipton, J. S., & Spelke, E. S. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Development, 76*, 978–988. http://dx.doi.org/10.1111/j.1467-8624.2005.00891.x.

Lyons, I. M., Ansari, D., & Beilock, S. L. (2012). Symbolic estrangement: Evidence against a strong association between numerical symbols and the quantities they represent. *Journal of Experimental Psychology: General, 141*, 635–641. http://dx.doi.org/10.1037/a0027248.

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition, 121*(2), 256–261. http://doi.org/10.1016/j.cognition.2011.07.009.

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science*, 1–13. http://dx.doi.org/10.1111/desc.12152.

Matejko, A. A., & Ansari, D. (2016). Trajectories of symbolic and nonsymbolic magnitude processing in the first year of formal schooling. *PLoS ONE, 11*(3), e0149863. http://dx.doi.org/10.1371/journal.pone.0149863.

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). *Preschoolers' Precision of the Approximate Number System Predicts Later School Mathematics Performance, 6*, 1–8. http://dx.doi.org/10.1371/journal.pone.0023749.

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

Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2016). How approximate and exact number skills are related to each other across development: A review. *Developmental Review, 39*, 1–15. http://dx.doi.org/10.1016/j.dr.2014.11.001.

Neter, J., Wasserman, W., & Kutner, M. H. (1985). *Applied linear statistical models*. Homewood, IL: Irwin.

Noël, M.-P., & Rousselle, L. (2011). Developmental changes in the profiles of Dyscalculia: An explanation based on a double exact-and-approximate number representation model. *Frontiers in Human Neuroscience, 5*(December), 165. http://dx.doi.org/10.3389/fnhum.2011.00165.

Nuerk, H.-C., Kaufmann, L., Zoppoth, S., & Willmes, K. (2004). On the development of the mental number line: More, less, or never holistic with increasing age? *Developmental Psychology, 40*, 1199–1211. http://dx.doi.org/10.1037/0012-1649.40.6.1199.

Passolunghi, M. C., & Lanfranchi, S. (2012). Domain-specific and domain-general precursors of mathematical achievement: A longitudinal study from kindergarten to first grade. *British Journal of Educational Psychology, 82*, 42–63.

Piazza, M., & Izard, V. (2009). How humans count: Numerosity and the parietal cortex. *The Neuroscientist : A Review Journal Bringing Neurobiology, Neurology and Psychiatry, 15*, 261–273. http://dx.doi.org/10.1177/1073858409333073.

Pureza, J. R., Jacobsen, G. M., Oliveira, R. G., & Fonseca, R. P. (2011). Relationships between executive functions tasks in late adulthood. *Psychology & Neuroscience, 4*, 369–376. http://dx.doi.org/10.3922/j.psns.2011.3.010.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20*, 110–122. http://dx.doi.org/10.1016/j.lindif.2009.10.005.

Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology, 91*(2), 137–157. http://dx.doi.org/10.1016/j.jecp.2005.01.004.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales. Section 2: The coloured progressive matrices*. Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

Repovs, G., & Baddeley, A. (2006). The multi-component model of working memory: explorations in experimental cognitive psychology. *Neuroscience, 139*, 5–21. http://dx.doi.org/10.1016/j.neuroscience.2005.12.061. doi:10.1111/j.2044-8279.2011.02039.x.

Reyna, V. F., & Brainerd, C. J. (2007). The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. *Learning and Individual Differences, 17*, 147–159. http://dx.doi.org/10.1016/j.lindif.2007.03.010.

Ritchie, S. (2015). *Intelligence: All that matters*. London: John Murray Learning.

Rousselle, L., & Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: a comparison of symbolic vs non-symbolic number magnitude processing. *Cognition, 102*(3), 361–395. http://doi.org/10.1016/j.cognition.2006.01.005.

Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2011). Association between basic numerical abilities and mathematics achievement. *The British Journal of Developmental Psychology, 30*, 344–357. http://dx.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, 67*, 37–41. http://dx.doi.org/10.1080/17470218.2013.803581.

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology, 114*, 418–431. http://dx.doi.org/10.1016/j.jecp.2012.10.012.

Sasanguie, D., & Reynvoet, B. (2014). Adults' arithmetic builds on fast and automatic processing of arabic digits: Evidence from an audiovisual matching paradigm. *PloS One, 9*(2), e87739. http://doi.org/10.1371/journal.pone.0087739.

Sasanguie, D., Van den Bussche, E., & Reynvoet, B. (2012). Predictors for mathematics Achievement? Evidence from a longitudinal study. *Mind, Brain, and Education, 6*, 119–128. http://dx.doi.org/10.1111/j.1751-228X.2012.01147.x.

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., et al. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*. http://dx.doi.org/10.1111/desc.12372.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development, 48*, 630–633. http://dx.doi.org/10.2307/1128664.

Simmons, F. R., Willis, C., & Adams, A.-M. (2012). Different components of working memory have different relationships with different mathematical skills. *Journal of Experimental Child Psychology, 111*, 139–155. http://dx.doi.org/10.1016/j.jecp.2011.08.011.

Starr, A., Libertus, M. E., & Brannon, E. M. (2013). *Number sense in infancy predicts mathematical abilities in childhood*. http://dx.doi.org/10.1073/pnas.1302751110. Proceedings of the National Academy of Sciences, 1–6.

Xenidou-Dervou, I., De Smedt, B., van der Schoot, M., & van Lieshout, E. C. D. M. (2013). Individual differences in kindergartessn math achievement: The integrative roles of approximation skills and working memory. *Learning and Individual Differences, 28*, 119–129. http://dx.doi.org/10.1016/j.lindif.2013.09.012.

Xenidou-Dervou, I., Gilmore, C., van der Schoot, M., & van Lieshout, E. C. D. M. (2015a). The developmental onset of symbolic approximation: Beyond nonsymbolic representations, the language of numbers matters. *Frontiers in Psychology, 6*(April), 1–13. http://dx.doi.org/10.3389/fpsyg.2015.00487.

Xenidou-Dervou, I., van Lieshout, E. C. D. M., & van der Schoot, M. (2014). Working memory in nonsymbolic approximate arithmetic processing: A dual-task study with preschoolers. *Cognitive Science, 38*, 101–127. http://dx.doi.org/10.1111/cogs.12053.

Xenidou-Dervou, I., van der Schoot, M., & van Lieshout, E. C. D. M. (2015b). Working memory and number line representations in single-digit addition: Approximate versus exact, nonsymbolic versus symbolic. *The Quarterly Journal of Experimental Psychology, 68*, 1–20. http://dx.doi.org/10.1080/17470218.2014.977303.