

Measuring the Approximate Number System in children: Exploring the  
relationships among different tasks.

Camilla Gilmore<sup>a</sup>, Nina Attridge<sup>a</sup>, Bert De Smedt<sup>b</sup> & Matthew Inglis<sup>a</sup>

<sup>a</sup> Mathematics Education Centre, Loughborough University, Leicestershire, LE11  
3TU, UK.

<sup>b</sup> Faculty of Psychology and Educational Sciences, University of Leuven, Leopold  
Vanderkelenstraat 32 - box 3765, 3000 Leuven, BELGIUM.

Corresponding author:

Camilla Gilmore

Tel: +44 (0)1509 228218

c.gilmore@lboro.ac.uk

7994 words

### Abstract

Research has demonstrated that children and adults have an *Approximate Number System* (ANS) which allows individuals to represent and manipulate the representations of the approximate number of items within a set. It has been suggested that individual differences in the precision of the ANS are related to individual differences in mathematics achievement. One difficulty with understanding the role of the ANS, however, is a lack of consistency across studies in tasks used to measure ANS performance. Researchers have used symbolic or nonsymbolic comparison and addition tasks with varying types and sizes of stimuli. Recent studies with adult participants have shown that performance on different ANS tasks is unrelated. Across two studies we demonstrate that, in contrast to adults, children's performance across different ANS tasks, such as symbolic and nonsymbolic comparison or approximate addition, is related. These findings suggest that there are differences across development in the extent to which performance on nonsymbolic and symbolic tasks reflects ANS precision.

Keywords: Mathematical cognition; Approximate Number System; Nonsymbolic arithmetic.

## 1. Introduction

In our everyday life we are surrounded by numbers and quantities: from the numbers on a page, to the items in a shopping basket and the coins in our pocket. Life becomes easier if we can efficiently remember, compare and order these quantities. We can choose the most numerous punnet of fruit, select the correct number of coins and join the shortest queue at the checkout. Children also deal with information about quantities in their play activities well before they start formal education. They may share sweets with their friends, compare the dots on two dice or count the pieces of a puzzle. Given the ubiquitous nature of numerical and quantity information, theorists have sought to understand how we represent and process numbers and magnitudes. In particular, researchers have explored whether individuals differ in how efficiently they can store and use numerical and quantity information, and how these differences arise. In this paper we explore how children represent and process information about quantities and whether there are consistent individual differences in their ability to do so.

Recently, psychologists have proposed the existence of an *Approximate Number System* (ANS), which is a cognitive system that allows us to represent and manipulate information about numbers and quantities (Cordes, Gelman, Gallistel & Whalen, 2001; Dehaene 1997; Feigenson, Dehaene & Spelke, 2004). An approximate representation of quantity is generated when we observe a set of items. According to the theory, quantity information is represented in an imprecise manner on a 'mental number line' where smaller quantities are represented more precisely than larger quantities. One version of the model proposes that quantity information about a set of  $n$  items will be represented

somewhere on the mental number line according to a normal distribution with a mean of  $n$  and standard deviation  $wn$  (Barth, La Mont, Lipton, Dehaene, Kanwisher & Spelke, 2006). Here the parameter  $w$  is known as the Weber fraction and varies across individuals so that the precision of people's ANS representations also varies. Individuals with a smaller  $w$  will tend to represent quantities more precisely than individuals with a larger  $w$ .

It has been suggested that the ANS is a universal system shared by infants, children, adults and animals (e.g., Barth, La Mont, Lipton & Spelke, 2005; Brannon, 2005; Dehaene, 1997; Pica, Lemer, Izard & Dehaene, 2004; Xu, Spelke & Goddard, 2005). ANS representations of quantity can be used to determine which of two sets of items has more elements, to order several sets of items according to the quantity or even to decide whether the sum of two sets is greater or less than a third set. In the lab this is typically explored by asking participants to compare dot arrays or sequences of tones. However, it is frequently argued that these abilities could serve a practical purpose in allowing an animal or human to determine, for example, which tree contained more fruit. Due to the imprecise nature of ANS representations, success in making comparisons of this sort will depend on the ratio between the two sets of items. It is more difficult to accurately determine which fruit tree, or array of dots, contains more elements if the ratio of the quantities is closer to one. The accuracy with which an individual can select the more numerous of two sets of items will thus depend both on the ratio between the sets and the size of the individual's  $w$ .

It has been suggested that ANS representations of quantity are also involved when individuals process symbolic numerical information, i.e. number words

and Arabic digits. Several authors have proposed that when exact symbolic representations of quantity are learned, they become mapped onto the pre-existing ANS representations (Le Corre & Carey, 2007; Noël & Rouselle, 2011; Wagner & Johnson, 2011). The timing and manner of this mapping is still under debate and may occur before (e.g. Wagner & Johnson, 2011) or after (e.g. Le Corre & Carey, 2007) children learn the cardinal meaning of symbolic numbers, or sometime later (e.g. Noël & Rouselle, 2011) and may or may not remain stable through life (Lyons, Ansari & Beilock, 2012). Nevertheless, there is evidence that by school age imprecise ANS representations are connected to symbolic representations and continue to influence performance on numerical tasks, even when they involve symbolic numerical representations. For example, participants show a ratio effect even when they are asked to choose the larger of two Arabic digits (e.g. Holloway & Ansari, 2009; Moyer & Landauer, 1967; Sekuler & Mierkiewicz, 1977). There is some evidence that as well as individual differences in the precision of the ANS representations themselves, there may be individual differences in the mapping between symbolic and nonsymbolic (ANS) representations. For example, Mundy and Gilmore (2009) found that individual differences in mapping skill accounted for variance in mathematical ability over and above the variance accounted for by nonsymbolic and symbolic comparison tasks.

Evidence for the ANS has come from studies showing that, within certain ratio limits, adults, children, infants and animals can identify the larger of two sets of items with above chance accuracy and show ratio effects on performance. For example, Barth et al. (2006) showed adult participants two dot arrays presented sequentially on the screen and asked them to select the more

numerous. The number of dots in each array ranged from 9 to 63 and the arrays were presented too quickly for the participants to be able to count them.

Participants' accuracy on this task was above chance levels, and varied according to the ratio between the quantities presented: participants were more accurate when the ratio between the quantities was 0.7 than when it was 0.8. Further studies have shown that 3 year-old children can also perform above chance on similar tasks and again show the characteristic ratio effect on performance (e.g., Libertus, Feigenson & Halberda, 2011). Infants as young as 6 months old can also detect changes in quantity in a habituation paradigm (Xu et al., 2005).

Much of the research exploring the ANS, particularly with children, has focused on discovering the types of ANS tasks that children can solve. So, for example, studies have revealed that children can compare, order, add, subtract and possibly even multiply and divide nonsymbolic quantities (typically dot arrays) with above chance accuracy (e.g. Gilmore, McCarthy & Spelke, 2007; McCrink & Spelke, 2010). As a result of this effort to explore the limits of the ANS, there has been little exploration of what factors contribute to individual differences in children's success on these tasks. However, the few studies to consider individual level analysis have revealed that there are wide variations in children's success with these problems. For example, McCrink & Spelke (2010) explored children's ability to perform nonsymbolic multiplication in a simple animated task involving dot arrays. Although group level analyses suggested that children were able to perform above chance levels on a task involving multiplication by 4, individual level analyses indicated that in fact only 9 out of 16 participants were performing above chance. A consideration of how children's ability to solve ANS tasks varies and the extent to which individual

differences are consistent across different ANS tasks is therefore overdue (Gilmore, 2013).

One area in which there has been a focus on individual differences concerns exploration of the relationship between ANS performance and mathematics achievement. Given that symbolic representations of number are mapped onto ANS representations, it has been proposed that the ANS may play a role in learning and performing mathematics and therefore individual differences in performance on ANS tasks will be related to individual differences in performance on mathematics tests. Data from a number of studies have supported this hypothesis (e.g. Bonny & Lourenco, 2013; Desoete, Ceulemans, De Weerd, & Pieters, 2012; Halberda, Mazocco & Feigenson 2008; Halberda, Ly, Wilmer, Naiman & Germine, 2012, Libertus et al., 2011; Libertus, Odic & Halberda, 2012; Libertus, Feigenson & Halberda, 2013; Lourenco, Bonny, Fernandez & Rao, 2012; Mazocco, Feigenson & Halberda, 2011) although other studies have failed to (e.g. Castronovo & Göbel, 2012; Holloway & Ansari, 2009; Iuclano, Tang, Hall & Butterworth, 2008; Kolkman, Kroesbergen & Leseman, 2013; Price, Palmer, Battista & Ansari, 2012; Sasanguie, Van den Bussche & Reynvoet, 2012; Sasanguie, Göbel, Moll, Smets & Reynvoet, 2013; Vanbinst, Ghesquière & De Smedt, 2012). The presence of this relationship may depend on the age of the participants and the nature of the tasks and measures involved (e.g. Inglis, Attridge, Batchelor & Gilmore 2011; Mundy & Gilmore, 2009; see De Smedt, Noël, Gilmore & Ansari, 2013 or Gilmore, 2013 for reviews).

Studies exploring this relationship have used a range of tasks to measure ANS performance (e.g. nonsymbolic or symbolic comparison, nonsymbolic addition) involving either small (1-9) or larger quantities (up to 70) and measures of

mathematics performance have included standardized mathematics tests, curriculum-based measures or simple arithmetic tasks. For example, Halberda et al. (2008) found that ANS performance in adolescence was related to mathematics ability in childhood, Libertus et al. (2011) found that 2-6-year old children's performance on a nonsymbolic comparison task was related to performance on a standardized mathematics test; Gilmore, McCarthy and Spelke (2010) found that 5-6-year old children's performance on a nonsymbolic addition task was related to performance on a school mathematics test; and Lyons and Beilock (2011) found that adults' performance on a nonsymbolic comparison task was related to their score on a mental arithmetic test. Alongside these studies, research on dyscalculia, a specific learning disorder with persistent difficulties in mathematics achievement, has shown that these children perform more poorly on nonsymbolic comparison tasks compared to typically developing controls (e.g., Piazza et al., 2010). However, several studies have failed to find a link between ANS performance and mathematics achievement, or found that this link is dependent on the age of participants, task or measure involved. For example, Inglis et al. (2011) found that nonsymbolic comparison performance was related to mathematics achievement in children aged 7-9- years but not for adult participants; Holloway and Ansari (2009) found that performance on a symbolic comparison task, but not a nonsymbolic comparison task was related to mathematics achievement in 7- to 9-year-old children; Rousselle and Noël (2007) and De Smedt and Gilmore (2011) observed that children with dyscalculia showed only impairments in symbolic but not nonsymbolic comparison tasks; Iuculano et al. (2008) found no relationships between nonsymbolic comparison, addition or subtraction and arithmetic skills;

and Price et al. (2012) found that adults' performance on three different variants of a nonsymbolic comparison task was unrelated to standardized maths performance.

One difficulty with interpreting the findings from these sets of studies is the range of tasks and stimuli that have been used. It is currently unknown whether tasks involving different operations, stimuli or number ranges give equivalent measures of ANS performance. If the ability to compare Arabic numerals up to 9 is unrelated to the ability to add large collections of dots, then it is unsurprising that there are varying results when comparing performance on just one of these tasks with mathematics achievement. However, few studies to date have used multiple ANS tasks within the same experiment, especially with child participants.

Some evidence exists from studies with adult participants to suggest that performance on different versions of ANS tasks is not related. Gilmore, Attridge and Inglis (2011) found low and non-significant correlations between adults' performance, indexed by accuracy or estimates of  $w$ , on a nonsymbolic comparison task and a nonsymbolic addition task even though the tasks involved the same number ranges and stimuli. However, adults' performance on a nonsymbolic comparison task involving large numerosity dot arrays was related to performance on a small-numerosity nonsymbolic comparison task, despite the use of different number ranges and stimuli. Performance on symbolic and nonsymbolic versions of the large numerosity comparison and addition tasks was unrelated although performances on symbolic and nonsymbolic versions of the small numerosity comparison task were significantly correlated. In contrast, Maloney, Risko, Preston, Ansari and Fugelsang (2010) found that performance,

indexed by the numerical distance effect, on nonsymbolic and symbolic versions of a small-numerosity comparison task was unrelated. Gebuis and van der Smagt (2011) also found that adult participants' performance on a nonsymbolic comparison task was unrelated to performance on an equivalent nonsymbolic detection task which involved the same stimuli and number range. Finally, Price et al. (2012) found that the strength of the relationships among three variants of a nonsymbolic comparison task depended on the measure used. There were low to moderate correlations between the tasks when the numerical distance effect was used as a measure of performance, but correlations were somewhat higher for *w* scores.

These studies suggest that different ANS tasks do not give equivalent measures of performance, at least in adults. Few studies have given multiple ANS tasks to children and so it is currently unknown whether individual differences in children's performance on different tasks are correlated. The little evidence that exists is mixed. A number of studies have failed to find correlations between performance on nonsymbolic and symbolic comparison tasks, using either accuracy or distance effect measures (Desoete et al., 2010; Holloway & Ansari, 2009; Lonneman, Linkersdörfer, Hasselhorn & Lindberg, 2011; Vanbinst et al., 2012). All of these tasks only examined performance for a small number range (maximum 9). Iuculano et al. (2008) did find correlations between performance on large-numerosity nonsymbolic comparison, addition and subtraction tasks using an efficiency measure of performance that incorporated accuracy and RT information, and Holloway and Ansari (2009) found a correlation between mean RT on nonsymbolic and symbolic versions of a small numerosity comparison task. However as both of these measures incorporated reaction times, it is

possible that the correlations simply reflected differences in processing speed across participants. It therefore remains to be established whether children's accuracy across a range of symbolic and nonsymbolic ANS tasks is related.

In this paper we explore whether a range of ANS tasks give related measures of children's performance and whether children's scores on these tasks can be assumed to index a single underlying system, namely the ANS. In the first study we examine 5- to 11-year-old children's performance on three different nonsymbolic tasks and in the second study we explore performance on nonsymbolic and symbolic versions of two different ANS tasks with children aged 5- to 7-years. We based all of the experimental tasks on examples in the existing literature.

## 2. Study 1

### 2.1 Method

#### 2.1.1 Participants

Participants were 74 children (33 male) who attended the University of Nottingham "Summer Scientist Week" event ([www.summerscientist.org](http://www.summerscientist.org)). Ages ranged from 5 years 0 months to 11 years 8 months ( $M = 7.6$  years). Children attending this event spent half a day at the university taking part in research studies as well as games and activities. Parents attended with their children and gave informed consent for all studies.

#### 2.1.2 Materials and procedure

Children took part in two 20-minute testing sessions during their visit. In one session the British Picture Vocabulary Scale (BPVS-II, Dunn, Dunn, Whetton & Burley, 1997) was administered and in the other session children completed three nonsymbolic tasks. The BPVS is a test of receptive vocabulary for children

aged 3 to 15 years. The experimenter reads a word and the participant is required to select which of four images matches the word. The test continues until the participant has answered 6 items incorrectly. The test produces raw scores and standardized scores ( $M = 100$ ,  $SD = 15$ ) and has good reliability (corrected split-half reliability = 0.86). The nonsymbolic tasks were presented on a laptop and children recorded their response using a button box. The order of presentation of the three nonsymbolic tasks was counterbalanced and roughly half of the children completed the computer tasks prior to the standardized test.

#### 2.1.2.1 Small-numerosity comparison

In this task children were shown two arrays of squares presented simultaneously on the screen. Their task was to select the more numerous array. The number of squares in each array ranged from 1 to 9, and the ratio between the arrays varied from 0.14 to 0.88. The stimuli were created in such a way that continuous quantity variables such as area and density of the squares could not be reliably used to select the correct array. The trials and stimuli used were identical to those used by Holloway and Ansari (2009). In each of 72 experimental trials children saw a fixation point followed by presentation of the stimuli for 1000 ms. Children could respond either when the stimuli were visible on the screen, or afterwards. The task was presented as a game in which the arrays showed how many sweets two cartoon characters had. Children were asked to select which of the two characters “has more”.

#### 2.1.2.2 Large-numerosity comparison

The large-numerosity comparison task was equivalent to the small-numerosity comparison task except in the stimuli involved. Children saw two arrays of dots presented simultaneously on the screen and their task was to

select the more numerous. The number of dots in each array ranged from 5 to 22 and in half of the trials the ratio between the numerosities of the arrays was 0.5 and in the remainder it was 0.7. To control for continuous quantity variables, the stimuli were created according to the method devised by Pica et al. (2004). In each of 64 experimental trials children saw a fixation point followed by presentation of the stimuli for 1500 ms. Children could respond either when the stimuli were visible on the screen, or afterwards. Children were asked to select which of two characters “has more”.

### 2.1.2.3 Approximate addition

In the approximate addition task children were required to add two red dot arrays and compare the result to a blue dot array. The addition task trials were matched to the large-numerosity comparison task trials so that for each of the comparison trials, one of the numerosities was split into two addends to create a new addition trial. The sum of the addends therefore ranged from 5 to 22 and in half of the trials the ratio between the sum totals and comparison number was 0.5 and in the remainder it was 0.7. In each of 64 experimental trials children saw a fixation point followed by an array of red dots presented for 1000 ms, this was followed by a fixation point then a second array of red dots presented for 1000 ms, a fixation point and an array of blue dots for 1000 ms and finally the question “Who has more?”.

## 2.2 Results

Initially, performance on each task was examined to verify that the expected ratio effects were observed (see Table 1 for performance on each task). Following this, the relationships among tasks were explored. Accuracy on the ANS tasks was used as the primary measure of performance. A range of

measures has been used in past studies (accuracy, RT,  $w$  scores, numerical distance effects). We chose accuracy because  $w$  estimates can be difficult to calculate for child participants (cf. Mazzocco et al., 2011), particularly where performance is close to chance, and RT data are also difficult to interpret if accuracy is close to chance. For performance that is significantly above chance,  $w$  scores and accuracies are very highly correlated (Inglis & Gilmore, submitted). For all analyses, Greenhouse-Geisser corrections were applied where sphericity assumptions were not met.

Split-half reliability for each of the tasks was calculated using the Spearman-Brown adjustment. Reliability of all the tasks was good (small-numerosity comparison = .86, large-numerosity comparison = .88, approximate addition = .87).

### 2.2.1 Task performance

Children's accuracy on the small-numerosity comparison task was examined with a one-way ANOVA (ratio: 0.25, 0.49, 0.75). The effect of ratio was significant ( $F(1.8, 125.1) = 161.1, p < .001, \eta_p^2 = .70$ ) and showed a significant linear trend ( $F(1,69) = 223.8, p < .001, \eta_p^2 = .76$ ). This demonstrates that, as expected, children's performance on this task was affected by the ratio between the to-be-compared items<sup>1</sup>.

Children's accuracy on the approximate addition and large-numerosity comparison tasks were examined with a 2 (task: addition, comparison)  $\times$  2 (ratio: 0.5, 0.7) repeated-measures ANOVA. As expected there was a significant ratio effect ( $F(1,67) = 60.2, p < .001, \eta_p^2 = .47$ ), children were more accurate at the 0.5 ratio ( $M = .79$ ) than the 0.7 ratio ( $M = .71$ ). There was also a main effect of

task ( $F(1,67) = 26.2, p < .001, \eta_p^2 = .28$ ) as children were more accurate at the addition task ( $M = .78$ ) than the comparison task ( $M = .71$ )<sup>2</sup>. The interaction between task and ratio did not reach significance ( $F(1,67) = 2.4, p = .123$ ), indicating that the size of the ratio effect was equivalent for the two tasks.

### 2.2.2 Relationships among tasks

To explore relationships amongst the tasks a series of zero-order and partial correlations were conducted (see Figure 1). Children's age and score on the BPVS were used to control for maturation and general verbal abilities, respectively. Accuracy on the approximate addition task was significantly correlated with performance on the large-numerosity comparison task ( $r = .59, p < .001$ ) and the small-numerosity comparison task ( $r = .44, p < .001$ ). When controlling for age and BPVS score these correlations remained significant ( $r = .53, p < .001$  and  $r = .39, p = .002$  respectively). Accuracy on the large-numerosity and small-numerosity comparison tasks was also significantly correlated ( $r = .58, p < .001$ ) and this relationship remained after controlling for age and general verbal abilities ( $r = .51, p < .001$ )<sup>3</sup>.

For each task there were a number of children who did not perform significantly above chance level (50%). To ascertain whether the correlations between task performance result from the poor performance of these children on each of the tasks, the correlations were re-run including only children who performed significantly above chance. There were 69 participants (93%) who performed significantly above chance for small-numerosity comparison, 51 participants (69%) for large-numerosity comparison and 66 participants (89%) for approximate addition. When only above-chance performance was considered, accuracy on the approximate addition task was significantly

correlated with performance on the large-numerosity comparison task ( $r = .43, p = .002$ ; controlling for age & BPVS  $r = .32, p = .041$ ) and the small-numerosity comparison task ( $r = .47, p < .001$ ; controlling for age & BPVS  $r = .41, p = .008$ ). Accuracy on the large-numerosity and small-numerosity comparison tasks was also significantly correlated ( $r = .55, p < .001$ ; controlling for age & BPVS  $r = .45, p = .003$ ). Performance across the different tasks was therefore significantly correlated when either the full or restricted samples were considered.

To determine whether performance on the ANS tasks represented the same underlying construct, accuracy on all three ANS tasks as well as BPVS score for all participants was entered into a principal components analysis (PCA) with varimax rotation. Bartlett's test ( $\chi^2 = 48.1, p < .001$ ) and the Kaiser-Meyer-Olkin statistic (0.67) both suggested that the data was suitable for PCA. Inspection of the scree plot and eigenvalues indicated that the data could best be described with two components, which together explained 77% of the variance (Component 1: 52%, Component 2: 26%). Component 1 had high loadings for all three ANS tasks (addition .810, large-numerosity comparison .861, small-numerosity comparison .805, BPVS -.028) indicating that the ANS tasks were measuring the same underlying construct. Component 2 had high loadings for verbal ability but low loadings for the ANS tasks (addition .131, large-numerosity comparison -.084, small-numerosity comparison -.113, BPVS .991). This analysis clearly demonstrates that all three ANS tasks appear to measure the same underlying construct, and that this is not the same as verbal ability.

To explore developmental trends in the data, the relationship between task performance and age was explored. There was a significant correlation between age and performance on each of the nonsymbolic tasks (addition  $r = .29,$

$p = .015$ ; large-numerosity comparison  $r = .37, p < .001$ ; small-numerosity comparison  $r = .43, p < .001$ ). To explore whether the relationships among tasks varied with age, residuals were calculated after regressing each nonsymbolic task onto the other nonsymbolic tasks and the correlation between the absolute residuals and age was examined. Age was not significantly correlated with the absolute residuals from the relationship between addition and large-numerosity comparison ( $r = -.11, p = .358$ ), addition and small-numerosity comparison ( $r = .16, p = .185$ ) or large- and small-numerosity comparison ( $r = .01, p = .909$ ). Thus there was no evidence that the relationships among the nonsymbolic tasks varied across the age range examined here.

### 2.3 Discussion

These findings indicate that children's performance on these tasks is moderately well correlated. The PCA indicates that performance on these tasks appears to be driven by the same underlying process. This contrasts with the evidence from adults, for whom performance on tasks similar to these showed very low correlations (Gilmore et al., 2011). While this might seem to suggest that the relationships among these tasks have a developmental trajectory, we did not observe such an effect across the age-range examined here.

Study 1 found a relationship in performance on ANS tasks in a sample of children with a wide age range (5 to 11 years) but similar backgrounds. A stronger case for a relationship in the ability to solve these tasks would be demonstrated by finding a relationship in a large sample of children with a narrow age range but who differed in other ways (e.g. SES, educational experience). A second study was therefore conducted to test the relationship between ANS tasks in a sample which included children aged 5 to 7 years who

were from differing backgrounds. These children came from varying educational environments as they were recruited from 19 different schools across two countries. The schools they were recruited from were situated in areas of varying socio-economic status. The study also included children with varying levels of mathematical ability and included children specifically recruited due to low achievement in mathematics.

Study 1 included only nonsymbolic versions of common ANS tasks and found that performance across these tasks was correlated in a sample of primary-school-aged children of a wide age range, which is contrary to previous findings from adults (Gilmore et al., 2011). Previous studies with adults have also found surprisingly low correlations between performance on symbolic and nonsymbolic versions of the same ANS task (Gilmore et al., 2011; Maloney et al., 2010). Given that our findings on the relationships between nonsymbolic tasks in children were different from previous findings with adults, the second study also included symbolic versions of ANS tasks, that were matched to the nonsymbolic tasks, to see whether the relationship between symbolic and nonsymbolic versions of the same tasks in children would also differ from previous findings with adults.

### 3. Study 2

#### 3.1 Method

##### 3.1.1 Participants

Participants were 140 children (47 male) aged between 5 years 6 months and 7 years 11 months ( $M = 6.5$  years). There were three groups of participants. One group included 46 children tested in primary schools in England (mean age = 6.1 years). They were recruited from 8 schools in mid- to low- socio-economic

status areas and were selected on the basis of low-achievement in mathematics: class teachers selected children who had below-average levels of achievement. The second group included 50 children tested in primary schools in Belgium (mean age = 6.6 years). They were recruited from 11 schools in mid- to high-socio-economic status areas and were selected on the basis of low-achievement in mathematics. All children scored below the 25<sup>th</sup> percentile on a curriculum-based standardized general mathematics test (Math Up to 10, Dudal, 1999). The final group included 44 children from the same primary schools in Belgium (mean age = 6.7 years). These children all had typical mathematics achievement, they scored above the 35<sup>th</sup> percentile on the same screening measure. Parents of all children gave consent for the study.

### 3.1.2 Materials and Procedure

Children were all tested in individual sessions in their school. Children completed two testing sessions on subsequent days. In one session they completed the nonsymbolic tasks and in the second session they completed the symbolic tasks. The tasks were presented on a laptop and children recorded their response using a standard keyboard attached to the laptop. In each session they also completed a small number of additional tasks that are not reported here, for further details please see De Smedt and Gilmore (2011). In this study the BPVS measure of verbal IQ was not included as there is no standardised Dutch version.

#### 3.1.2.1 Nonsymbolic and symbolic small numerosity comparison

These tasks involved the same procedure and trials but differed only in the format of the stimuli. Children were shown two quantities presented side-by-side on the screen and were asked to decide which quantity was the more

numerous. In the nonsymbolic task the quantities were presented as arrays of circles, and in the symbolic version of the task the quantities were presented as Arabic digits. Trials consisted of all combinations of the quantities 1 to 9. The ratio between the quantities varied from 0.11 to 0.89. In each of 72 experimental trials children saw a fixation point accompanied by a beep followed by the two quantities presented until response. In the symbolic version of the task, children were instructed to select the largest number. In the nonsymbolic version of the task, children were instructed to select the array which had the largest number of dots. As in Study 1, nonsymbolic arrays were created using the method devised by Pica et al. (2004) to control for continuous quantity variables.

#### 3.1.2.2 Nonsymbolic and symbolic addition

These two tasks involved the same trials and similar procedure but differed in the nature of the stimuli. Children watched an animated sequence in which they saw two quantities appear on one side of the screen and a comparison quantity appear on the other side. They were asked to add the two quantities on the left and decide whether the total was more or less than the quantity on the right. In the nonsymbolic version of the task, quantities were presented as arrays of dots. Children saw one dot array appear on the left of the screen and move behind an occluder, then a second dot array appeared on the left of the screen and moved behind the same occluder. Finally a third dot array appeared on the right side of the screen and remained visible. The experimenter narrated the sequence e.g. "Daniel has some marbles and puts them in a box, and he puts some more into his box. Now all Daniel's marbles are in his box. Look. Paul also has some marbles. Can you tell me who has more?". In the symbolic version of the tasks, quantities were presented as coloured boxes labelled with

Arabic digits. Children saw one box appear on the left of the screen and remained visible, then a second box appeared on the left of the screen. Finally a third box appeared on the right side of the screen and remained visible. The experimenter narrated the sequence e.g. “Tommy has five candies, and he gets five more. James has 50 candies. Who has more?”. There were 24 experimental trials for each version of the task. Numbers involved in the task ranged from 5 to 58 and the approximate ratio between the sum total and the comparison number was 0.8, 0.7 or 0.6.

### 3. Results

Initially, performance on each task was examined to verify that the expected ratio effects were observed (see Table 2 for performance on each task). Following this, the relationships among tasks were explored. For all analyses, Greenhouse-Geisser corrections were applied where sphericity assumptions were not met.

Split-half reliability was again calculated for each of the tasks using the Spearman-Brown adjustment. Reliability of the comparison tasks was good (nonsymbolic comparison = .83, symbolic comparison = .85) but reliability of the approximate addition tasks was poor (nonsymbolic addition = .38, symbolic addition = .53).

#### 3.1 Task performance

Children’s accuracy on the approximate addition tasks was explored using a 2 (version: nonsymbolic, symbolic)  $\times$  3 (ratio: 0.6, 0.7, 0.8) ANOVA. There was a significant main effect of version ( $F(1,127)=24.2, p < .001, \eta_p^2 = .16$ ) with accuracy higher for the symbolic ( $M = .68$ ) than the nonsymbolic version ( $M = .62$ ) of the task. As expected there was a significant main effect ( $F(1.9,242.0) =$

67.8,  $p < .001$ ,  $\eta_p^2 = .35$ ) and linear trend ( $F(1,127) = 109.5$ ,  $p < .001$ ,  $\eta_p^2 = .46$ ) of ratio. There was also an interaction between version and ratio ( $F(2,254) = 7.1$ ,  $p = .001$ ,  $\eta_p^2 = .05$ ). Performance on the two versions of the small-numerosity comparison task was examined using a 2 (version: nonsymbolic, symbolic)  $\times$  3 (ratio: 0.23, 0.5, 0.77) ANOVA. There was a significant main effect of version ( $F(1,137) = 5.2$ ,  $p = .025$ ,  $\eta_p^2 = .04$ ) with accuracy higher for the symbolic ( $M = .91$ ) than the nonsymbolic version ( $M = .89$ ) of the task. There was also the expected main effect ( $F(1.8,247.3) = 394.5$ ,  $p < .001$ ,  $\eta_p^2 = .74$ ) for ratio but the interaction between version and ratio ( $F(1.7,236.3) = 2.4$ ,  $p = .099$ ) did not reach significance.

### 3.2 Relationships among tasks

To explore the relationships among task performance, two sets of correlations were undertaken (see Figure 2). First, performance on the nonsymbolic tasks was considered. When all participants were included, there was a significant correlation between accuracy on the nonsymbolic small-numerosity comparison task and the nonsymbolic addition task ( $r = .23$ ,  $p = .009$ ; controlling for age  $r = .19$ ,  $p = .042$ ). Next, the relationship between nonsymbolic and symbolic versions of each task was examined. For the whole sample, there was a significant correlation between nonsymbolic and symbolic versions of the small-numerosity comparison task ( $r = .49$ ,  $p < .001$ ; controlling for age  $r = .45$ ,  $p < .001$ ), and between nonsymbolic and symbolic versions of the approximate addition task ( $r = .28$ ,  $p = .002$ ; controlling for age  $r = .24$ ,  $p = .008$ ).

As in Study 1, the full sample included many children who did not perform significantly above chance level (50%) on one or more tasks. Therefore

these correlations were re-run with participants who performed at chance on a task removed from the analysis involving that task. This left 133 participants (95%) for nonsymbolic small-numerosity comparison, 134 participants (96%) for symbolic small-numerosity comparison, 66 participants for nonsymbolic approximate addition (47%) and 91 participants (65%) for symbolic approximate addition. With this restricted sample there was again a significant correlation between accuracy on the nonsymbolic small-numerosity comparison task and the nonsymbolic addition task ( $r = .26, p = .036$ ; controlling for age  $r = .36, p = .013$ ). There was also a significant correlation between nonsymbolic and symbolic versions of the small-numerosity comparison task ( $r = .49, p < .001$ ; controlling for age  $r = .52, p < .001$ ), however the correlation between nonsymbolic and symbolic versions of the approximate addition task was not significant ( $r = .20, p = .161$ ; controlling for age  $r = .13, p = .386$ ).

### 3.3 Discussion

In Study 2, with participants who had a broader range of educational experiences and mathematical abilities, we replicated the key findings of Study 1. We found a small but significant correlation between performance on a nonsymbolic comparison task and a nonsymbolic addition task. Furthermore, we showed that performance on nonsymbolic and symbolic versions of both the comparison task and addition task were correlated. In this study we did not include a measure of general verbal abilities, as we did in Study 1. Therefore we are unable to demonstrate here that the correlations among ANS tasks remain after controlling for general abilities. However, given that performance on ANS tasks were unrelated to general verbal abilities in Study 1, it is unlikely that general verbal abilities account for the present findings.

#### 4. General discussion

In this paper we investigated for the first time children's performance on a range of nonsymbolic and symbolic tasks designed to recruit the Approximate Number System (ANS). We explored whether children's performance across these tasks was correlated and reflected the performance of a single underlying system. In two studies we found that children's performance of nonsymbolic addition was significantly correlated with their performance of nonsymbolic comparison. Furthermore we found significant correlations across performance of nonsymbolic and symbolic comparison tasks. These findings contrast with previous research with adult participants. Here we consider why the patterns of performance for children may differ from those observed with adult participants and explore what our findings reveal about the way that the ANS is recruited to solve tasks across development.

In both of our studies we found significant correlations between children's performance on nonsymbolic addition and comparison tasks ( $r_s \approx .2$  to  $.5$ ). This contrasts with the findings of Gilmore et al. (2011) who found that adults' performance on similar nonsymbolic addition and comparison tasks was not significantly correlated ( $r_s \approx .05$ ). There are at least three possible explanations of these differences: a) these types of ANS tasks may have better psychometric properties and show higher reliability for children than adults; b) the lack of a correlation in the adult sample may have resulted from a restricted range of abilities leading to insufficient variance in scores; or c) there are differences between children and adults in the extent to which performance on these tasks reflect ANS acuity vs. general processing demands. These possible explanations will be considered in turn.

One potential account of the differences between our findings and those involving adult participants is that these tasks have better psychometric properties with children than adults. If the tasks have poor internal and test-retest reliability with adult participants then low correlations among the tasks are unsurprising. On the whole split-half reliability estimates for the administered ANS tasks were adequate to good ( $r_s \approx .8$ ). The exception to this is the approximate addition tasks in Study 2 which had poor split-half reliability. The latter two tasks only included 24 trials involving three different comparison ratios, which may have reduced the internal reliability. Previous explorations of the psychometric properties of these types of tasks also do not suggest that these tasks have poorer psychometric properties with adults than children. In fact some evidence suggests that these types of tasks have better psychometric properties with adult participants than children. Gilmore et al. (2011) found that both symbolic and nonsymbolic versions of addition and comparison tasks had satisfactory split-half reliability for accuracy with adult participants, Maloney et al. (2010) also found that nonsymbolic comparison tasks had adequate reliability with adult participants, for both accuracy and RT, although reliability for symbolic tasks was lower and Sasanguie, Defever, Van den Bussche and Reynvoet (2011) found RT-based performance measures from a nonsymbolic comparison task were reliable. In contrast, estimates for reliability of ANS tasks with child participants appear to be lower. Schneider, Grabner and Paetsch (2009) found that children's performance indexed by RT on a symbolic comparison task had low test-retest reliability across an interval of 6 days. Using a very similar nonsymbolic comparison task to that used here in Study 1, Inglis and Gilmore (submitted) found that accuracy measures had high immediate test-

retest reliability with both adult and child participants, but that one-week test-retest reliability was higher with adult than child participants. Thus, the difference between our findings and previous research with adult participants cannot be explained by recourse to better psychometric properties of the tests with child vs. adult participants.

On the other hand, the contrast between the current study findings and previous research with adult participants may have arisen because the adult study included university student participants, who may have a restricted range of abilities, compared to child studies, which included children with a broader range of abilities. Research exploring the characteristics of the ANS in adult participants have typically involved university samples (although see Halberda et al. (2012) for an exception). It would be informative to include a wider range of participants in future studies to explore whether restriction of range problems may have accounted for the current contrasting findings.

Finally, an alternative explanation of stronger correlations across nonsymbolic addition and comparison tasks for children than adults is developmental differences in the extent to which performance on these tasks reflect ANS acuity vs. domain-general demands of the task. Evidence is beginning to emerge regarding the domain-general demands of ANS tasks such as those used here. Xenidou-Dervou, Van Lieshout & Van der Schoot (in press) investigated the working memory demands of approximate addition tasks and found that the central executive was particularly implicated. Similarly, Fuhs and McNeil (2013) highlighted the inhibitory control demands of nonsymbolic comparison tasks. Both of these studies involved young participants in preschool or kindergarten and it is clear that the domain-general demands of these tasks

will be substantial for young children. If the addition and comparison tasks present substantial executive function demands for children, then scores on both of these tasks may reflect domain-general skills more than ANS acuity. This might lead to the observed correlation in performance across these tasks. For adult participants, however, it is possible that the domain-general demands may be less and differ across tasks and this might reduce the correlations between task performance. There is currently no data on the domain-general demands of these tasks for adult participants, but this is an important avenue to explore.

A greater appreciation of the domain-general demands of these ANS tasks can help to shed light on the nature of the relationship between performance on ANS tasks and mathematics achievement. As previously discussed, there is mixed evidence concerning the relationship between performance on nonsymbolic comparison tasks and achievement in mathematics. The existence of such a relationship appears to depend on the specific characteristics of the tasks and the age of the participants. Since these are factors that will impact on the domain-general demands of the task, it is interesting to consider the extent to which these factors may account for the relationship with mathematics performance. Fuhs and McNeil (2013) recently demonstrated that inhibition skills may account for the relationship between nonsymbolic comparison performance and mathematics achievement in a sample of young children from low income backgrounds. The fact that correlations among ANS tasks and correlations between ANS tasks and mathematics are both stronger in children than adults suggests that a common explanation may be warranted. Domain-general demands of the tasks is one candidate explanation that requires further investigation.

As highlighted in the introduction, this is the first study to explore relationships between nonsymbolic addition and comparison tasks in children. However, previous research has investigated the relationship between nonsymbolic and symbolic versions of a small numerosity comparison task. In Study 2 we found that the correlation between accuracy on these tasks was 0.5. This contrasts with previous studies that have found non-significant correlations between performance on nonsymbolic and symbolic comparison tasks (e.g. Holloway & Ansari, 2009; Lonnemann et al., 2011; Sasanguie et al., 2013; Vanbinst et al., 2012). We believe that this difference is most likely to be due to the use of different performance measures. Studies that have failed to find a correlation across these tasks have used ratio effect or reaction time measures whilst we used accuracy. It has been shown in previous research that different measures of performance from the same task are not correlated. Price et al. (2012) found no correlation between  $w$  estimates and ratio effects on a nonsymbolic comparison task with adult participants. Inglis and Gilmore (submitted) found low correlations between ratio effects and accuracy on a nonsymbolic comparison task with adults and children. Moreover, ratio effect measures showed lower reliability than accuracy or  $w$  estimates. Ratio effect measures index changes in a participant's performance across different ratios but do not capture absolute levels of performance (so two participants may have equivalent ratio effects but very different levels of performance). In contrast, accuracy will reflect overall levels of performance.

In both of the studies reported here a large number of children failed to perform above chance-levels on one or more of the tasks. In Study 1 only 57% of children were above chance on all of the tasks and for Study 2 this figure was

only 34%. These figures highlight that although, as a group, children are able to recruit their ANS to solve these types of problems, there are large individual differences in their ability to do so. Children's poor performance on these tasks may have arisen from imprecise ANS representations but may also stem from difficulties in focusing on the numerical rather than visual features of the arrays, or the working memory demands of the task (particularly the addition tasks). Some evidence suggests that children's performance on nonsymbolic tasks is affected by the visual characteristics of the arrays (e.g. Soltész, Szucs & Szucs, 2010), although the mechanisms involved are, at present, unclear. In general, there has been little research exploring how domain-general factors such as these influence performance on ANS tasks. In order to understand why there are such great individual differences in performance on ANS tasks, we need to better understand what cognitive resources and visual information are involved in solving these types of problems. This may, in turn, help us to understand the mixed findings concerning the relationship between performance on ANS tasks and formal symbolic mathematics.

In conclusion, our findings have revealed that several nonsymbolic tasks designed to recruit the ANS do indeed index the same system in children, contrary to previous research with adults. This suggests that the tasks may be measuring different things in adults and in children. However, there were large individual differences in children's ability to solve the tasks indicating that a focus on individual differences would help us better understand what influences mathematics performance across development.

## Notes

<sup>1</sup> Similar effects were found on the small-numerosity comparison tasks when performance was analysed according to numerical distance, rather than ratio.

<sup>2</sup>Significantly better performance on addition than comparison tasks was predicted by a model of the ANS proposed by Barth et al. (2006). However, in their study Barth and colleagues found that adults' performance did not show this pattern and they modified the model accordingly. Our findings show that children's performance does match the original model, and this highlights a potential difference between adults and children on the pattern of performance on these tasks.

<sup>3</sup>The small- and large-numerosity comparison tasks had overlapping set sizes (small 1 – 9, large 5 – 22). When performance is examined on subsets of these tasks with no overlapping set sizes (small 1 – 7, large 8 – 22; 44 trials per task) then the correlation between performance on each of the tasks remains ( $r = .251$ ,  $p = .039$ ).

### Acknowledgements

Study 2 was funded by a small research grant from the British Academy awarded to CG and BDS. CG is supported by a British Academy Postdoctoral Fellowship and MI is supported by a Royal Society Worshipful Company of Actuaries Education Research Fellowship. We thank all the families who attended the University of Nottingham Summer Scientist Week 2011 ([www.summerscientist.org](http://www.summerscientist.org)) and all other children and their families who participated in these studies.

## References

- 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.
- Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 14116–14121.
- 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*, 375–388.
- Brannon, E. M. (2005). *What animals know about numbers*. In J. I. D. Campbell (Ed.). *Handbook of mathematical cognition*. New York: Psychology Press.
- Castronovo, J., & Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PLoS ONE*, *7*, e33832.
- 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*, 698–707.
- 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.
- De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*. 48 – 55.
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

- Desoete, A., Ceulemans, A., De Weerd, F., & Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? Findings from a longitudinal study. *British Journal of Educational Psychology, 82*, 64–81.
- Dudal, P. (1999). *Rekenbasis tot 10 [Math up to 10]*. Torhout: PMS Centrum 2.
- Dunn, L., Dunn, L., Whetton, C. & Burley, J. (1997). *The British Picture Vocabulary Scale II*. Windsor: NFER-Nelson.
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*, 307–314.
- Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: contributions of inhibitory control. *Developmental Science, 16*, 136–148.
- Gebuis, T., & van der Smagt, M. J. (2011). False Approximations of the Approximate Number System? *PLoS ONE, 6*, e25405.
- Gilmore, C. (2013). Approximate arithmetic in childhood. In A. Dowker & R. Cohen Kadosh (Eds). *The Oxford Handbook of Mathematical Cognition*. Oxford: OUP.
- 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*, 589–591.
- 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.
- Halberda, J., Mazocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*, 665–668.

- 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, 109*, 11116–1120.
- 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.
- 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.
- Inglis, M. & Gilmore, C. (submitted). Indexing the Approximate Number System  
*Manuscript submitted for publication.*
- Iuculano, T., Tang, J., Hall, C. W. B., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science, 11*, 669–680.
- Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction, 25*, 95–103.
- 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*, 395–438.
- 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.

- 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.
- Libertus, M. E., Odic, D., & Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychologica, 141*, 373-379.
- Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., & Lindberg, S. (2011). Symbolic and non-symbolic distance effects in children and their connection with arithmetic skills. *Journal of Neurolinguistics, 24*, 583-591.
- Lourenco, S. F., Bonny, J. W., Fernandez, E. P., & Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences, 109*, 18737-18742.
- Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition, 121*, 256-261.
- 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.
- Maloney, E. A., Risko, E. F., Preston, F., Ansari, D., & Fugelsang, J. (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica, 134*, 154-161.
- Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE, 6*, e23749.

- McCrink, K., & Spelke, E. S. (2010). Core multiplication in childhood. *Cognition*, *116*, 204–216.
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*, 1519–1520.
- Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology*, *103*, 490–502.
- 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*.
- Piazza, M., Facoetti, A., Trussardi, A., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*, 33–41.
- Pica, P., Lemer, C., V. Izard, & Dehaene, S. (2004). Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science*, *306*, 499–503.
- Price, G. R., Palmer, D., Battista, C., & Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychologica*, *140*, 50–57.
- Rousselle, L., & Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: a comparison of symbolic vs. non-symbolic number magnitude. *Cognition*, *102*, 361-395.
- Sasanguie, D., Defever, E., Van den Bussche, E., & Reynvoet, B. (2011). The reliability of and the relation between non-symbolic numerical distance effects. *Acta Psychologica*, *136*, 73-80.

- 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.
- 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.
- Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: their interrelations in grades 5 and 6. *Journal of Educational Psychology*, *101*, 359–372.
- Sekuler, R. & Mierkiewicz, D. (1977). Children's Judgments of Numerical Inequality. *Child Development*, *48*, 630-633.
- Soltész, F., Szucs, D., & Szucs, L. (2010). Relationships between magnitude representation, counting and memory in 4- to 7-year-old children: A developmental study. *Behavioral and Brain Functions*, *6*.
- Vanbinst, K., Ghesquière, P., & De Smedt, B. (2012). Numerical Magnitude Representations and Individual Differences in Children's Arithmetic Strategy Use. *Mind, Brain, and Education* *6*, 129–136.
- Wagner, J. B., & Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition*, *119*, 10–22.
- Xenidou-Dervou, I., Van Lieshout, E.C.D.M. & Van der Schoot, M. (2013). Working memory in nonsymbolic approximate arithmetic processing: A dual-task study with pre-schoolers. *Cognitive Science*. DOI: 10.1111/cogs.12053

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants.

*Developmental Science*, 8, 88–101.

Table 1: Performance on tasks in Study 1.

Task	<i>M</i>	<i>SD</i>	Range
Small-numerosity comparison (accuracy)	0.87	0.09	0.54 - 0.97
Large-numerosity comparison (accuracy)	0.71	0.13	0.41 - 1.00
Approximate addition (accuracy)	0.78	0.11	0.50 - 0.97
BPVS (standard score*)	108.67	11.44	79 - 133

\* BPVS standardization sample  $M = 100$ ,  $SD = 15$  (Dunn et al., 1997)

Table 2: Accuracy on nonsymbolic and symbolic tasks in Study 2.

Task	<i>M</i>	<i>SD</i>	Range
Nonsymbolic small-numerosity comparison	0.86	0.08	0.61 - 0.99
Symbolic small-numerosity comparison	0.88	0.09	0.51 - 1.00
Nonsymbolic approximate addition	0.61	0.10	0.42 - 0.92
Symbolic approximate addition	0.67	0.12	0.42 - 0.96

### Figure captions

Figure 1: Relationships amongst accuracy on nonsymbolic tasks from Study 1: (a) approximate addition and large-numerosity comparison; (b) approximate addition and small-numerosity comparison; (c) small- and large-numerosity comparison.

Figure 2: Relationships amongst accuracy on tasks from Study 2: (a) nonsymbolic addition and nonsymbolic small-numerosity comparison; (b) nonsymbolic and symbolic comparison; (c) nonsymbolic and symbolic approximate addition.





