

Symbolic Number Skills Predict Growth in Nonsymbolic Number Skills in Kindergarteners

Ian M. Lyons
University of Western Ontario

Stephanie Bugden
University of Pennsylvania

Samuel Zheng and Stefanie De Jesus
Toronto District School Board, North York, Ontario, Canada

Daniel Ansari
University of Western Ontario

There is currently considerable discussion about the relative influences of evolutionary and cultural factors in the development of early numerical skills. In particular, there has been substantial debate and study of the relationship between approximate, nonverbal (approximate magnitude system [AMS]) and exact, symbolic (symbolic number system [SNS]) representations of number. Here we examined several hypotheses concerning whether, in the earliest stages of formal education, AMS abilities predict growth in SNS abilities, or the other way around. In addition to tasks involving symbolic (Arabic numerals) and nonsymbolic (dot arrays) number comparisons, we also tested children's ability to translate between the 2 systems (i.e., mixed-format comparison). Our data included a sample of 539 kindergarten children ($M = 5.17$ years, $SD = .29$), with AMS, SNS, and mixed-comparison skills assessed at the beginning and end of the academic year. In this way, we provide, to the best of our knowledge, the most comprehensive test to date of the direction of influence between the AMS and SNS in early formal schooling. Results were more consistent with the view that SNS abilities at the beginning of kindergarten lay the foundation for improvement in both AMS abilities and the ability to translate between the 2 systems. It is important to note that we found no evidence to support the reverse. We conclude that, once one acquires a basic grasp of exact number symbols, it is this understanding of exact number (and perhaps repeated practice therewith) that facilitates growth in the AMS. Though the precise mechanism remains to be understood, these data challenge the widely held view that the AMS scaffolds the acquisition of the SNS.

Keywords: numerical cognition, numerical representation, number symbols, approximate magnitudes, numeracy and development

In the search for the origins of numerical abilities, there has been a debate regarding the extent to which basic, possibly innate, skills that humans share with nonhuman animals serve as the foundation for more complex skills that are taught more deliberately via cultural practices and, in particular, via formal education. Conversely, it is of considerable interest whether cultural inputs may shape and refine more evolutionarily basic abilities.

In the domain of numerical cognition and mathematics education, a major current debate concerns the interplay between the ability to discriminate between approximate, nonverbal magnitudes (such as arrays of dots) and the ability to represent numbers in exact, symbolic form (e.g., as with Indo-Arabic numerals). The former, approximate and nonverbal capacity is shared across many species (for reviews, see Agrillo & Beran, 2013; Nieder & Dehaene, 2009; Pahl, Si, & Zhang, 2013) and, at least in basic form, is thought to be present from birth in humans (e.g., Lipton & Spelke, 2003; Xu & Spelke, 2000). Recently, it has become apparent that humans and other species can do more than just compare nonverbal magnitudes; they can even perform simple, approximate arithmetic (such as sums and ratios; e.g., Brannon, Wusthoff, Gallistel, & Gibbon, 2001; Capaldi & Miller, 1988; Matthews, Lewis, & Hubbard, 2016; McCrink & Spelke, 2010, 2016). Such abilities are thought to be underpinned by what is often referred to as the approximate number system, or ANS. There is some debate over the extent to which these approximate magnitudes are strictly numerical (e.g., Leibovich & Henik, 2013); hence, for present purposes, we adopt the broader term: approximate magnitude system (AMS).

Number symbols are cultural in origin, because they rely on arbitrary social conventions to determine their forms and the basic

This article was published Online First November 20, 2017.

Ian M. Lyons, Department of Psychology, University of Western Ontario; Stephanie Bugden, Department of Psychology, University of Pennsylvania; Samuel Zheng and Stefanie De Jesus, Research and Information Services, Toronto District School Board, North York, Ontario, Canada; Daniel Ansari, Department of Psychology, University of Western Ontario.

This research was supported by funding from the Canadian Institutes of Health Research, The National Sciences and Engineering Research Council of Canada, Canada Research Chairs program to Daniel Ansari, as well as from the Toronto District School Board, and a Banting Postdoctoral Fellowship Award to Ian Lyons.

Correspondence concerning this article should be addressed to Ian M. Lyons, who is now at the Department of Psychology, Georgetown University, White-Gravenor Hall 302D, Washington, DC 20057. E-mail: ian.lyons@georgetown.edu

rules that govern how they are to be mathematically manipulated (Zhang & Norman, 1995). We refer here to the ability to represent and manipulate number symbols as the symbolic number system, or SNS. Over the course of development, children go from viewing symbols as meaningless shapes to having a rich understanding of their meaning. A key question is how children develop an understanding of the meaning of number symbols and to what representations these symbols become linked during this ontogenetic process.

One view that has garnered considerable attention is that the AMS plays a crucial role as the foundation for and scaffold of the SNS (Dehaene, 1997, 2008; Feigenson, Dehaene, & Spelke, 2004; Feigenson, Libertus, & Halberda, 2013; Gallistel & Gelman, 2000; Piazza, 2010). In other words, the evolutionarily basic capacity, in the form of the AMS, is thought to provide the critical foundation for the culturally acquired capacity (the SNS). This is an intuitive hypothesis, capitalizing on the broader notion that cultural skills must at some level coopt extant, evolutionarily ancient, neural structures (Dehaene & Cohen, 2007). Moreover, many of the basic operations that form the basis of the SNS—such as relative quantity (greater–lesser), ordinality, and even simple arithmetic—are available, at least in approximate form, to the AMS (e.g., Brannon et al., 2001; Capaldi & Miller, 1988; Matthews et al., 2016; McCrink & Spelke, 2010, 2016). Furthermore, numerous studies have shown that the precision of an individual’s AMS is predictive of SNS abilities (for a review and meta-analysis, see Chen & Li, 2014). There is also work to suggest that arithmetic training using approximate, nonsymbolic magnitudes (arrays of dots) leads to improvement in symbolic calculation scores (Hyde, Khanum, & Spelke, 2014; Park & Brannon, 2013, 2014). Taken together, these results strongly suggest that one’s early formal understanding of the SNS may be bootstrapped from the informal AMS. A straightforward prediction is that children who begin kindergarten with stronger AMS skills should be best positioned to acquire early formal SNS skills during the school year: Children with strong AMS skills at the beginning of the year should show the strongest growth in SNS skills over the course of the year.

An alternative view is that once one acquires a basic grasp of exact number symbols, it is this exact understanding of number (and repeated practice therewith) that facilitates growth in the AMS. In other words, approximate, nonverbal quantities increasingly come to be understood in symbolic terms (Mix, 2008; Mix, Huttenlocher, & Levine, 1996; Mix, Huttenlocher, & Levine, 2002; Mussolin, Nys, Leybaert, & Content, 2016). From this perspective, it is the culturally acquired SNS that refines the evolutionarily more basic capacity (the AMS). Though perhaps somewhat counterintuitive, there is evidence to support the notion that cultural inputs may have an influence on the AMS. For example, Piazza, Pica, Izard, Spelke, and Dehaene (2013) examined adults and children among the Mundurucú (an indigene group in the Amazon) and found that those with access to education showed higher AMS acuity—specifically, higher precision when comparing arrays of dots. No difference as a function of education was found on a task requiring participants to determine which of two disks was larger in area. These results suggest that educational inputs may increase the precision of the evolutionarily more basic capacity to process approximate magnitudes. That said, one cannot rule out the possibility that those with higher AMS acuity were predisposed to seek out educational opportunities. Moreover, it is

not clear precisely what educational inputs (e.g., math education in particular or socialization aspects of the educational environment) may have been responsible for the observed difference in AMS acuity. Mussolin, Nys, Content, and Leybaert (2014) recently provided evidence suggesting that symbolic numerical skills in particular predict improvement in AMS acuity. Mussolin et al. examined 57 Belgian preschoolers (ages 3–4) at two time points (roughly seven months apart). They showed that scores on a symbolic numerical battery at Time 1 significantly predicted growth in AMS acuity (i.e., AMS acuity at Time 2, after controlling for AMS acuity at Time 1). It is interesting that the reverse relation was both nonsignificant and significantly smaller (i.e., $SNS1 \sim \Delta AMS > AMS1 \sim \Delta SNS$), suggesting the direction of influence runs specifically from the SNS to the AMS and not the other way around (for a similar result in a sample of 30 first graders, see also Matejko & Ansari, 2016).

Linking Symbolic and Nonsymbolic Numerical Processing

Whether early numerical influences flow primarily from the AMS to the SNS or the other way around, both accounts require some mechanism for linking the two systems to transmit this influence. From the AMS \rightarrow SNS perspective, one such suggestion is that approximate representations of quantity in fact serve as the semantic content of their symbolic counterparts (e.g., Dehaene, 2008; Feigenson et al., 2004, 2013; Piazza, 2010; Piazza, Pinel, Le Bihan, & Dehaene, 2007; Verguts & Fias, 2004). This view presupposes a strong link between the AMS and SNS more or less from the time one begins to acquire number symbols. On the other hand, from the SNS \rightarrow AMS perspective, one need not presuppose a direct link between the SNS and AMS even early in development. Hence, an important additional piece of the debate is the need for and possible development of an ability to translate between symbolic and approximate numerical representations (Brankaer, Ghesquière, & De Smedt, 2014; Mundy & Gilmore, 2009).

One way to probe this translational ability is to ask participants to match or compare quantities across (symbolic and nonsymbolic) formats. In a study with adults, Lyons, Ansari, and Beilock (2012) demonstrated that the ability to compare a symbolic (numeral) stimulus with a nonsymbolic (dot array) stimulus is nontrivial: Mixed-format comparisons took significantly longer than did either symbolic or nonsymbolic same-format comparisons. The cost of mixing numerals and dots was also significantly higher than the cost of mixing numerals and number–words, suggesting the key distinction is not necessarily visual format but whether the stimuli point to symbolic (SNS) or nonsymbolic (AMS) representations. Recent neural evidence has also provided evidence consistent with this distinction (Bulthé, De Smedt, & Op de Beeck, 2014; Bulthé, De Smedt, & Op de Beeck, 2015; Damarla, Cherkassky, & Just, 2016; Damarla & Just, 2013; Lyons, Ansari, & Beilock, 2015). In other tasks that force one to translate between AMS stimuli and verbal symbols (number–words), one finds systematic biases (Crollen, Castronovo, & Seron, 2011; Izard & Dehaene, 2008). On the other hand, each of these studies—including Lyons et al. (2012)—used literate adults as participants. Thus, the distinction–separation between the AMS and SNS may arise only over time, after considerable educational experience has perhaps shifted the focus onto symbolic representations of number. Indeed, Lyons et

al. strongly implied that their results were likely specific to literate adults and suggested that the AMS and SNS may be strongly linked early in development and become “estranged” only later in development. This proposal is more in keeping with the AMS → SNS perspective because it assumes a central role for the AMS in shaping early SNS representation. From this perspective—in particular, assuming the early meaning of number symbols is derived directly from their nonsymbolic counterparts (e.g., Dehaene, 2008; Feigenson et al., 2004; Piazza et al., 2007; Verguts & Fias, 2004)—one would expect the cost of mixing numerals and dots to be minimal in younger children, with this cost increasing over developmental time as the two systems (AMS and SNS) become increasingly distinct.

An alternative view is that the AMS and SNS are actually formed more or less independently of one another (Carey, 2011; Le Corre & Carey, 2007; Lyons & Ansari, 2015). This would imply a high cost of mixing symbolic and nonsymbolic inputs early in development. It is intriguing that the SNS → AMS view outlined earlier would predict a *reduction* in the cost of mixed-format comparisons over development. This is because one’s understanding of approximate, nonverbal quantities is increasingly informed by one’s knowledge of and experience with exact symbolic representations of numbers (Mix, 2008; Mix et al., 1996; Shusterman, Slusser, Halberda, & Odic, 2016), thereby *increasing* the link between the AMS and the SNS over developmental time.

Here it is also important to note that the AMS → SNS view and SNS → AMS view differ in terms of what is more likely to lead to improvement in the ability to translate between symbolic and nonsymbolic quantities. From the former (AMS → SNS) perspective, an individual with high AMS acuity should have less trouble affixing approximate, nonsymbolic quantities to exact symbolic representations: Early AMS skills should predict growth in symbolic–nonsymbolic mixing skills. From the latter (SNS → AMS) perspective, nonsymbolic quantities should increasingly be understood in terms of their symbolic counterparts: Early SNS skills should predict growth in symbolic–nonsymbolic mixing skills. Hence, it is crucial to assess the degree to which either early AMS or early SMS (or both) skills predict developmental change in how well children translate between the two systems.

Current Study

In the current study, we empirically probe these two accounts of early symbolic number development—along with the hypotheses each raises—in a single, large, longitudinal study focusing on over 500 children over the course of kindergarten (i.e., in the fall and then the spring). Hence, we focused on the early stages of formal education to address the broader questions (a) To what extent are the AMS and SNS linked at the outset of formal education? and (b) Does the AMS shape the SNS, or the other way around? In sum, here we provide the most comprehensive directional test to date of the early relation between the AMS and the SNS.

We examined this question by testing several complementary hypotheses (summarized in Table 1).¹ First, we examined the cost of mixing between formats in a manner similar to that in Lyons et al. (2012) by computing the degree to which performance on a mixed-comparison task was worse than that on dot- and numeral-comparison tasks (in particular, the critical difference is between mixed comparison and whichever shows worse performance—

dots or numerals). Specifically, because the AMS is thought to form the foundation of the SNS, the AMS → SNS view predicts a minimal early (in the fall) cost of mixing formats (Hypothesis 1a; see Table 1 for a summary of hypotheses) that either increases or stays constant over time (from fall to spring; Hypothesis 2a). The SNS → AMS view assumes the two systems are initially distinct and so predicts a large early (in the fall) mixing cost (Hypothesis 1b) that, due to reinterpretation of the AMS in terms of the SNS, lessens over time (from fall to spring; Hypothesis 2b).

Next, we examined whether early (fall) AMS or SNS ability is a better predictor of growth (change from fall to spring) in the ability to translate between the two systems. Under the assumption that the AMS provides the critical foundation for the culturally acquired SNS, the AMS → SNS view predicts that dot comparison in the fall will be a better predictor of mixed-comparison growth (Hypothesis 3a); the SNS → AMS view predicts that numeral comparison in the fall will be a better predictor of mixed-comparison growth (Hypothesis 3b).

Finally, we examined whether early (fall) AMS ability predicts growth in SNS ability (change from fall to spring), the other way around, or both. The AMS → SNS view holds that dot comparison in the fall will be a better predictor of numeral-comparison growth (Hypothesis 4a); the SNS → AMS view holds that numeral comparison in the fall will be a better predictor of dot-comparison growth (Hypothesis 4b). The relation could be bidirectional (Hypothesis 4c, as suggested by Feigenson et al., 2013; Mussolin et al., 2016). Note also that, should results from testing the preceding hypotheses provide evidence that the AMS and the SNS are more likely to be distinct systems even in kindergarteners, it is still important to test Hypothesis 4. Two distinct systems can still influence one another; thus, one could posit a modified version of the AMS → SNS view in which the AMS influences SNS development without being part of the same underlying system.

Method

Participants

Participants were 613 children in Senior Kindergarten kindergarten² from 36 schools in the greater Toronto area (all schools are part of the Toronto District School Board, TDSB, whose students comprise one of the largest and most diverse school districts in Canada). Of these, 539 children completed all three critical number-comparison tasks at both time points; subsequent analyses proceeded with $N = 539$ (241 female). Mean age at the time of the first testing session was 5.17 years ($SD = .29$, range = 4.67–5.77SD =). Sixty-five children were not born in Canada. Socio-

¹ Our measure of AMS ability was a standard dot-comparison task; SNS ability was measured via a numeral comparison task; and the ability to translate between the two systems was measured via a mixed-format comparison task.

² Note that in several Canadian provinces, Kindergarten is divided into ‘Junior’ and ‘Senior’ Kindergarten. The former is akin to what is often referred to elsewhere as ‘preschool’; this typically takes place when children are about 4 years old and is often relatively informal in overall structure. Senior Kindergarten is perhaps more closely related to what is referred to as Kindergarten in other areas. Instruction at this phase, while not as strictly structured as first grade, nevertheless begins to emphasize many basic formal concepts in mathematics and other areas.

Table 1
Hypotheses and Predictions

Theoretical view and hypothesis	Prediction	Explanation
AMS → SNS		
1a	MC1 = DC1	No cost of mixing formats
2a	(DC1–MC1) < (DC2–MC2)	Mixing cost increases over time
3a	DC1 → ΔMC	Early AMS predicts growth in mixing formats
4a	DC1 → ΔNC	Early AMS predicts growth in SNS
SNS → AMS		
1b	MC1 < DC1	Cost of mixing formats already present
2b	(DC1–MC1) > (DC2–MC2)	Mixing cost decreases over time
3b	NC1 → ΔMC	Early SNS predicts growth in mixing formats
4b	NC1 → ΔDC	Early SNS predicts growth in AMS
SNS ↔ AMS		
4c	(DC1 → ΔNC) and (NC1 → ΔDC)	Early AMS predicts growth in SNS and early SNS predicts growth in AMS

Note. AMS = approximate magnitude system; SNS = symbolic number system; MC = mixed comparison; DC = dot comparison; NC = numeral comparison; 1 = Time 1 (fall); 2 = Time 2 (spring); Δ = growth (Time 2 controlling for Time 1).

economic status (SES) was not available for individual children; however, it could be estimated for each school. Schools were classified as 0 = *Low-SES* (24%), 1 = *Medium-Low-SES* (31%), 2 = *Medium-High-SES* (34%), and 3 = *High-SES* (11%).

Procedure

All behavioral data were collected using paper and pencil Numeracy Screener booklets (complete versions of the screeners can be found in the online supplemental materials). Data were collected at two time points: fall of 2014 and spring of 2015.³ The average interval between time points was 191.84 days ($SD = 14.30$, range = 141–217). The same version of the booklets was used at each time point.

Data were collected in collaboration with teachers, early childhood educators (ECEs), and administrators in TDSB schools. Testing materials were approved by the University of Western Ontario's Non-Medical Research Ethics Board. The data reported here are part of a larger joint research project between the TDSB and the University of Western Ontario (UWO), which was approved by the TDSB's External Research Review Committee (ERRC). The TDSB's research department was authorized by the board to collect student personal information and assessment data for the purposes of the board's educational planning. For this joint research project, parents of participating students were informed by their respective schools that the assessment data were to be collected by the classroom educators and that the confidential student-level data collected were to be kept strictly within the TDSB's Research and Information Services. Only depersonalized data without any school or student identifiers were shared with external partners (including researchers at UWO).

All data collection was completed by teachers and early childhood educators from the classrooms where testing took place. Testing was conducted on a one-on-one basis with each student in a separate, quiet area, requiring approximately 15–20 min per student. Teachers and ECEs were provided with an in-service work day during which they were given explicit training on administering test booklets. Written instructions were also provided for each task in the booklet.

For each task, teachers went over a predefined set of instructions with the child. These instructions were printed in the booklet at the start of each task, along with other guidelines for administering the task. The teacher went over several examples with the child. They then explained to the child,

You should try to complete as many problems as you can. You have two minutes. Work as fast as you can without making too many mistakes. If you make a mistake, draw an X through the mistake and put a new line through the right answer.

Teachers then demonstrated how to correct a mistake. A corrected answer was counted as correct. Children were also instructed to complete the items for a given task in the order they were presented, without skipping items ("Make sure you don't skip any items"). Once the child was ready, the teacher started the timer, and the child turned the page to begin.

Comparison Tasks

Task booklets were based loosely on the design originally innovated by *Nosworthy, Bugden, Archibald, Evans, and Ansari (2013)*. Booklets contained six numeral tasks, though only three of these—the three numeral-comparison tasks, described in detail in the next sections—are of direct theoretical relevance to the questions and hypotheses outlined earlier. Children completed three comparison tasks: numeral comparison, dot comparison, and mixed comparison—always in that order. Each comparison task comprised 72 total items, with 12 items per page. Children were given 2 min to complete as many items as possible per task.

Numeral comparison (NC). Examples of the numeral-comparison task are shown in *Figure 1a*. Children were told, "In this task, your job is to decide which of the two numbers is bigger. Draw a line through the box with the number that means the most things." Numerals ranged from 1 to 9, with absolute numerical

³ The majority of children were tested in October (89%) and May (90%), though a few were tested in November (4%) or December (7%) in the fall or April (7%) or June (3%) in the spring. Median test dates were October 27 (fall) and May 8 (spring).

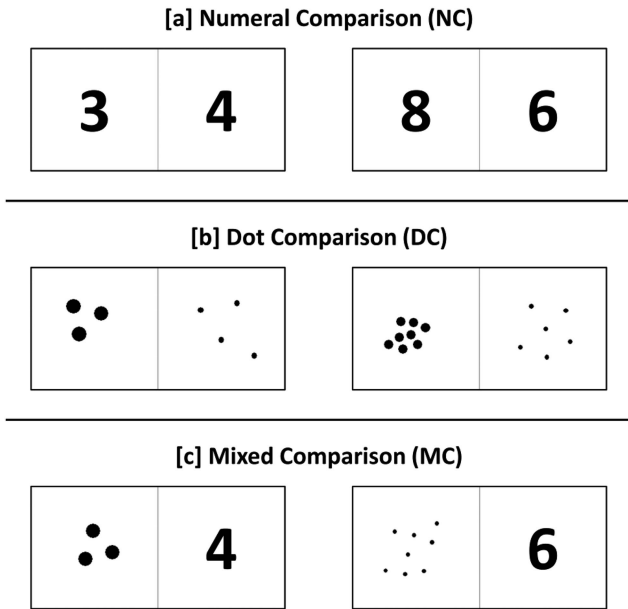


Figure 1. Examples of the numeral- (Panel a), dot- (Panel b), and mixed- (Panel c) comparison tasks.

distances $|n_1 - n_2|$ of 1 to 3 and ratios (minimum–maximum) from .250 to .889. Specifically, all 15 combinations of 1–9 with distances of 1 or 2 were included, along with three combinations with distance 3 ($\{1,4\}$, $\{3,6\}$, $\{6,9\}$). This yielded 18 possible combinations. Of these 18, nine were permuted such that the larger number was on the left, and the other nine were permuted such that the larger number was on the right. The nine trials were chosen such that the larger side was in no way related to numerical size, distance, or ratio. The next 18 trials were arranged in the opposite manner. The last 36 trials were determined in the same manner. Trial order was then pseudorandomized within each set of 18 trials such that, for any n th item in the sequence, average numerical ratio, size, and distance were equated across comparison tasks (numeral, dot, mixed). This final step ensured that if, for instance, a given child completed exactly 10 trials on each of the three comparisons' tasks, the ratios (or sizes or distances) encountered on each task would not have differed significantly across tasks (all p s > .20). In other words, comparing performance across tasks was not confounded with these numerical factors.

Dot comparison (DC). Examples of the dot-comparison task are shown in Figure 1b. Children were told, “In this task, your job is to decide which of two boxes contains more dots. Draw a line through the box that has the most dots in it.” Children were also instructed, “Don’t try to count the dots. Instead, just look at the dots and try your best to guess which side has more dots in it.” Numerosities and trial order were determined in the same manner as in the numeral-comparison task just described. In addition, two versions of a given permutation were created. In one version, dot area was positively correlated with numerosity, and overall contour length was negatively correlated with numerosity; in the other version, the opposite was true. On a given trial, the two parameters were thus in opposition; between trials, relying on any single parameter would have led to chance performance (Gebuis &

Reynvoet, 2012). Parameter version order was further pseudorandomized such that it was not informative of the correct answer within a given segment of trials.

Mixed comparison (MC). Examples of the mixed-comparison task are shown in Figure 1c. Children were told,

In this task, your job is to decide whether a number or a group of dots means more things. If the number means more things, draw a line through the number. If the dots mean more than the number, then draw a line through the dots.

As with dot comparisons, children were also instructed, “Don’t try to count the dots. Instead, just look at the dots and try your best to guess which side means more.” Numerosities and trial order were determined in the same manner as in the numeral-comparison task described earlier. In addition, which side contained the numeral and which side the dots was pseudorandomized such that it was not informative of the correct answer within a given segment of trials.

Scoring. Raw scores were computed as the net number of items correctly completed within the 2-min time limit. In a timed task such as this, it is crucial to adjust for guessing; randomly guessing on all 72 items would yield, on average, a raw score of 36. Hence, scores were adjusted for guessing using this standard adjustment:

$$A = C - \frac{I}{(P - 1)},$$

where A is the adjusted score, C is the number correct, I is the number incorrect, and P is the number of response options (Rowley & Traub, 1977). This method has the effect of adjusting a guessing strategy, on average, to 0. For instance, in a four-item multiple-choice exam (where each choice is equally probable), randomly guessing on 20 items would yield, on average, an unadjusted score (C in the equation) of 5. The adjusted score (A) would be

$$5 - \frac{15}{(4 - 1)} = 0.$$

In the current case, each item had two alternatives (left and right quantity, equally likely to be correct), so adjusted scores were effectively *correct* minus *incorrect* ($A = C - I$). That said, results were highly similar regardless of whether adjusted or raw scores were used, indicating that the presence of guessing strategies did not substantially influence the results. Mean adjusted scores are given in Figure 2.

Covariates

For all regression analyses, the following variables were included as control variables: age (years), sex, whether a child was born in Canada (0 or 1), school SES, percentage of days absent during the kindergarten school year ($M = 8.9\%$, $SD = 7.2$, range = 0–50.3),⁴ and testing interval (in days).⁵

⁴ Absentee rates were not available for three children. These children were assigned the average rate (8.9%). Results did not differ if these children were excluded.

⁵ One of the two testing dates was not recorded for 20 children (and hence testing interval could not be computed). These children were assigned the average interval (191.84 days). Results did not differ if these children were excluded.

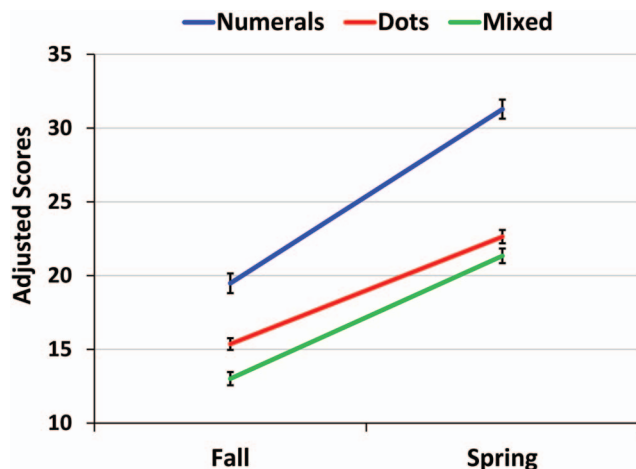


Figure 2. Mean performance (adjusted scores: Correct – Incorrect) for the three comparison tasks in the fall (Time 1) and spring (Time 2). Error bars indicate standard errors of the mean. See the online article for the color version of this figure.

In addition, we controlled for children who may not have understood a given task. Rather than exclude a child altogether for failing to perform above chance (adjusted score >0) on a single task, we instead created a dummy variable for each task, coded as 1 if a child's adjusted score ≤ 0 , or 0 if a child's adjusted score >0 . In any regression model that included a given task (as either an independent or a dependent variable), the corresponding dummy variable for that task was also included as a covariate of no interest. Note that these variables were necessarily (anti-)correlated with the variables of interest: actual performance (adjusted or otherwise). This makes this approach relatively conservative, indicating that we have perhaps slightly underestimated effect sizes (i.e., partial correlations). It is important to note, therefore, that all nonsignificant ($p \geq .05$) results remained so, whether or not these dummy variables were included.

Predicting Growth

To test Hypotheses 3 and 4 (regarding the direction of influence between the AMS and SNS), we examined which tasks at Time 1 uniquely predicted growth in which of the other tasks. As recommended by Castles and Coltheart (2004; see also a similar discussion in Sasanguie, Defever, Maertens, & Reynvoet, 2014), growth was assessed by predicting spring (Time 2) scores after controlling for fall (Time 1) scores. This is an estimate of growth because it removes the Time 1 variance from both the predictor and the outcome, meaning that any residual relation between the predictor and outcome is, by definition, specific to what is unique to the outcome at Time 2.⁶ Predicting Time 2 while controlling for Time 1 is also preferable because the within-task relation across time (e.g., DC Time 1 \sim DC Time 2) is interpretable. However, such relations (i.e., the circular arrows in Figure 3) are not, strictly speaking, predictions of growth, though, of importance, they are unique to that task, because the other tasks (e.g., NC and MC at Time 1) are of course included in the model. In this way, consistent with Mussolin et al. (2014), we computed three regression models, predicting growth in each of the three comparison tasks. For

example, to predict numeral-comparison growth, we adjusted the outcome scores on spring (Time 2) numeral comparison, and the main predictors of interest were fall (Time 1) numeral, dot, and mixed adjusted scores (along with all the relevant covariates, as noted earlier).

Reliability

Due to time constraints, we used timed tests for the three comparison tasks; interitem reliability (i.e., Cronbach's alpha) could not be computed over all 72 possible items in each task because not all children completed all 72 items (indeed, few did). Instead, we identified the maximum number of trials that at least two thirds of participants completed across all three tasks (we used Time 1 data here, because that was the first time children encountered all tasks). The first 20 trials for dot-, numeral-, and mixed-comparison tasks were completed by 70.3%, 68.8%, and 67.7% of participants, respectively. Limiting reliability estimates to just these participants and just the first 20 trials, we found reasonable to good reliability for all three tasks (dot comparison: $\alpha = .70$, numeral comparison: $\alpha = .83$, mixed comparison: $\alpha = .69$; results were similar if other arbitrary thresholds were adopted).

Though reliabilities were fairly comparable to one another, that for numeral comparison was a bit higher than for the other two tasks, which may have inflated regression results for this task. Indeed, differences in measurement reliability are known to pose a major potential confound for cross-lagged panel longitudinal models, such as the type we used here (Hamaker, Kuiper, & Grasman, 2015; Rogosa, 1980). To address this concern, we recomputed the partial correlations in Figure 3, but only after disattenuating the relation between each pair of critical variables for the reliabilities of the relevant variables (e.g., Murphy & Davidshofer, 2004). Our central conclusions from testing Hypotheses 3 and 4 remained unchanged (see Appendix C for complete results), though it is important to acknowledge that, if some alternative conception of the comparison measures' reliabilities were devised in the future, results might change.

Manipulation Check—Ratio Effects

An important assumption of the AMS view is that nonsymbolic magnitude processing reliably elicits ratio effects (as the ratio between the two quantities being compared approaches 1, performance diminishes). It was thus important to check that the dot-comparison task in particular reliably elicited ratio effects. We assessed this at Time 1 (when children were least familiar with the tasks) by examining whether accuracy decreased as ratio increased (ratio was computed here as minimum–maximum). For simplicity,

⁶ Note that this method is preferred over using change scores (e.g., Time 1–Time 2) for several reasons. First, change scores contain variance from both Time 1 and Time 2. If one found a correlation, for example, between numerals (NC) at Time 1 and dot (DC) change scores, it would not be clear whether this was due to DC variance at Time 1 or at Time 2 (or both). One can include DC scores at Time 1 as a covariate to remove this variance; however, this yields results identical to just predicting Time 2 after controlling for Time 1 (as we have done here). Second, a given score at Time 1 will necessarily be correlated with a change score based on that same variable (e.g., DC at Time 1 predicting DC change scores), making such results largely uninterpretable.

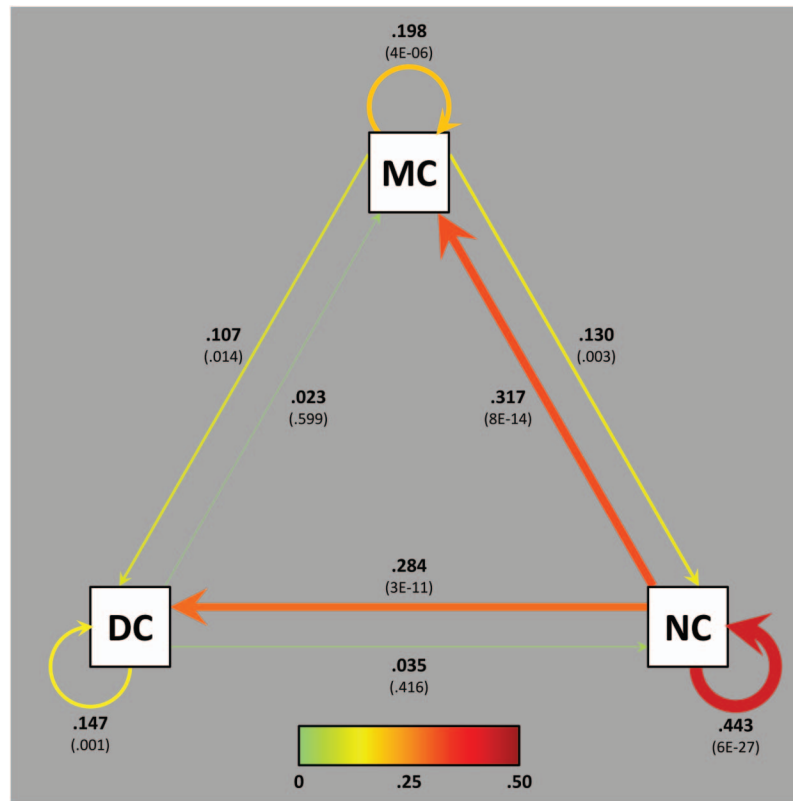


Figure 3. Shows results from the longitudinal models predicting growth in each task. An arrow from one task to another indicates the partial correlation between the originating task at Time 1 and the indicated task at Time 2, controlling for all other tasks at Time 1 (as well as the various covariates). For example, the arrow from NC to MC indicates the unique relation between Time 1 numeral comparison performance and growth in the mixed comparison task (because Time 1 mixed comparison was included in the model). Arrows are weighted and colored according to partial- r values (range: 0 to .5). Values in bold are actual partial- r values; values in parentheses are corresponding p -values. Circular arrows indicate the unique relation between a given task at Time 1 and itself at Time 2. Abbreviations: MC = mixed comparison; NC = numeral comparison; DC = dot comparison. See the online article for the color version of this figure.

we binned trials into far (ratio $\leq .6$, $M = .421$), mid ($.6 < \text{ratio} \leq .75$, $M = .710$), and near ($.75 < \text{ratio}$, $M = .839$). Overall accuracy clearly fell as a function of ratio in all three tasks: DC: far = 91.68% ($SE = .68$), mid = 77.50% ($SE = .92$), near = 64.51% ($SE = .72$); MC: far = 88.13% ($SE = .83$), mid = 73.31% ($SE = 1.00$), near = 65.63% ($SE = .83$); NC: far = 88.47% ($SE = .92$), mid = 81.99% ($SE = 1.01$), near = 78.00% ($SE = .94$). It is also important to see that ratio reliably predicted accuracy on an individual child level (Lyons, Nuerk, & Ansari, 2015). For each child, we computed the correlation (r value) between ratio and accuracy, where a ratio effect would be indicated by a strong negative correlation. Average r values for the three tasks were as follows: DC: mean $r = -.735$ ($SE = .016$), MC: mean $r = -.635$ ($SE = .021$), NC: mean $r = -.366$ ($SE = .024$). In sum, the dot- (and mixed-)comparison task reliably elicited ratio effects at the population and individual levels, indicating it was indeed measuring the AMS. Though the numeral-comparison task showed a ratio effect when averaging across all children, this was much less reliable for individual subjects (consistent with what was seen among children in Grades 1–6; Lyons, Nuerk, & Ansari, 2015).

Results

For all tests, effect sizes and exact p values are given for additional context; the significance threshold was $p < .05$. Raw data can be found here: <https://osf.io/uf2gb/>.

Developmental Changes and Task Differences

To test Hypotheses 1a and 1b and Hypotheses 2a and 2b (whether there is a cost of mixing formats at the outset of kindergarten, and how this changes over the course of the school year), we examined longitudinal changes in performance from fall to spring for the three numeral-comparison tasks. The AMS \rightarrow SNS view predicts that mixed-comparison performance will be no worse than dot-comparison performance in the fall (Hypothesis 1a); this difference should either remain constant or increase in the spring (Hypothesis 2a). Hypothesis 2 predicts that this difference should be significantly reduced at Time 2 (spring). The SNS \rightarrow AMS view predicts that mixed-comparison performance will be worse than dot-comparison performance in the fall (Hypothesis 1a); this difference should either remain constant or increase in the

spring (Hypothesis 2a). Hypothesis 2 predicts that this difference should be significantly reduced at Time 2 (spring).

Data were first entered into a 3 (task: numeral, dot, mixed) \times 2 (time: fall, spring) within-subject analysis of variance (results are summarized in Figure 2). The main effect of task was significant, $F(2, 1076) = 335.33, p = 6E-114, d = 1.58$. Performance was overall best on numeral comparisons, followed by dot comparisons, and performance was worst on mixed comparisons. Numeral-comparison scores were significantly higher than were dot-comparison scores at both time points: fall: $t(538) = 8.62, p = 7E-17, d = .74$; spring: $t(538) = 19.53, p = 1E-64, d = 1.68$. Dot-comparison scores were significantly higher than were mixed-comparison scores at both time points: fall: $t(538) = 7.10, p = 4E-12, d = .61$; spring: $t(538) = 4.04, p = 6E-05, d = .35$. Note that this latter result indicates that comparing between formats was more difficult than comparing within either symbolic or nonsymbolic formats. This result is thus more consistent with the view that symbolic and nonsymbolic representations of number form dissociable representation systems at the very outset of formal schooling, which is consistent with Hypothesis 1b but not 1a. In addition, the difference between dot and mixed comparison was significantly less at Time 2: $F(1, 538) = 6.41, p = .012, d = .22$, which is consistent with Hypothesis 2b but not 2a. In sum, tests of both Hypotheses 1 and 2 supported the SNS \rightarrow AMS view: There was a cost of mixing symbolic and nonsymbolic formats even at the outset of kindergarten, and this cost decreased over the course of the school year.

With respect to characterizing change from Time 1 to Time 2 as “growth,” it is crucial to note that the main effect of time was significant, $F(1, 1076) = 669.08, p = 2E-96, d = 2.23$. All three tasks showed significant improvement from fall to spring: NC: $t(538) = 23.73, p = 1E-85, d = 2.05$; DC: $t(538) = 17.60, p = 4E-55, d = 1.52$; MC: $t(538) = 19.59, p = 7E-65, d = 1.69$.

It is also important to note that the Task \times Time interaction was significant, $F(2, 1076) = 5.07, p = 2E-21, d = .61$. Longitudinal improvement was significantly⁷ greater for NC than either DC, $t(538) = 9.09, p = 2E-18, d = .78$, or MC, $t(538) = 6.91, p = 1E-11, d = .60$, and it was significantly greater for MC relative to DC, $t(538) = 2.53, p = .012, d = .22$ (equivalent to the F test performed to test Hypothesis 2 earlier). These differences in longitudinal improvement are notable in several respects. First, they indicate differential developmental change with respect to symbolic, nonsymbolic, and mixed number processing. Second, they argue against the notion that task improvements may have been

driven simply by familiarity with the tasks (which would predict merely a main effect of time).

Predicting Longitudinal Growth

In this section, we tested Hypotheses 3 and 4 (what the direction of influence between the AMS and the SNS is). According to the AMS \rightarrow SNS view, dot comparison at Time 1 should be a strong unique predictor of both mixed-comparison growth (Hypothesis 3a) and numeral-comparison growth (4a). According to the SNS \rightarrow AMS view, numeral comparison at Time 1 should be a strong unique predictor of both mixed-comparison growth (Hypothesis 3b) and dot-comparison growth (4b). Finally, it is of course possible to find a bidirectional influence on growth (Hypothesis 4c).

To test these hypotheses, we assessed the degree to which symbolic, nonsymbolic, and mixed numeral processing at the outset of formal education (i.e., at the beginning of kindergarten) uniquely predicted growth in one another over the course of the school year. Unique contributions were assessed via multiple regression, controlling not just for competing variables of interest but also for several covariates of no interest: age, sex, testing interval, absentee rates, school SES, birth location (in Canada or not), and chance performance (see the Method section). Growth was assessed by predicting spring (Time 2) scores after controlling for fall (Time 1) scores.

Model results are visualized in Figure 3 (zero-order correlations between comparison scores are given in Table 2). Arrows in Figure 3 indicate growth directionality. For instance, the arrow pointing from DC to NC in Figure 3 denotes the degree to which dot comparisons at Time 1 predict growth in numeral comparisons; the arrow pointing the opposite direction denotes the degree to which numeral comparisons at Time 1 predict growth in dot comparisons (partial r and p values are also given; full model results can be found in Appendix A).

From Figure 3, it is clear that early (Time 1) numeral-comparison performance was a strong unique predictor of growth in the other two tasks, dot comparison did not significantly predict growth in either of the other two tasks, and mixed comparison was a moderate predictor of growth in the other tasks. Time 1 numeral comparison was a significantly⁸ stronger unique predictor of growth in mixed comparison than Time 1 dot comparison (.317 vs. .023; $z = 4.95, p = 7E-07$), which is consistent with Hypothesis 3b but not 3a. Time 1 numeral comparison was also a significantly stronger unique predictor of dot-comparison growth than Time 1 mixed comparison (partial r s = .284 vs. .108; $z = 2.99, p = .003$).

From Figure 3, it can be seen that symbolic skills are the primary unique predictor of growth in all three tasks. Indeed, Time 1 numeral comparison was in fact a stronger unique predictor of Time 2 mixed comparison and dot comparison than either of those tasks at Time 1 were of themselves (numeral vs. mixed comparison: .317 vs. .198; $z = 2.07, p = .038$; numeral vs. dot comparison: .284 vs. .147; $z = 2.34, p = .019$). Notably, the reverse was not

Table 2
Zero-Order Correlations Between the Three Comparison Tasks at Times 1 and 2

Variable	NC1	DC1	MC1	NC2	DC2	MC2
1. NC1	—	.706	.738	.716	.590	.624
2. DC1	.706	—	.698	.548	.533	.512
3. MC1	.738	.698	—	.623	.526	.595
4. NC2	.716	.548	.623	—	.737	.790
5. DC2	.590	.533	.526	.737	—	.773
6. MC2	.624	.512	.595	.790	.773	—

Note. NC = numeral comparison; DC = dot comparison; MC = mixed comparison; 1 = Time 1 (fall); 2 = Time 2 (spring). All correlations were significant (all p s $\leq 2E-37$).

⁷ Here, this is reported as t tests between change scores (spring–fall), which is equivalent to the relevant 2×2 interaction term.

⁸ This was computed by comparing partial correlations using Fisher’s z tests. Fisher tests were chosen because, within a given model, partial r s are by definition independent (e.g., the partial relation between x_1 and y controls for x_2 , and vice versa).

true: Time 1 numeral comparison was a stronger predictor of itself at Time 2 than were Time 1 mixed comparison (.443 vs. .130; $z = 5.62$, $p = 2E-08$) and Time 1 dot comparison (.443 vs. .035; $z = 7.17$, $p = 8E-13$).

Moreover, Time 1 numeral comparison asymmetrically predicted growth in the other two comparison tasks. Time 1 numeral comparison was a stronger unique predictor of dot-comparison growth than was Time 1 dot comparison of numeral-comparison growth (partial $r_s = .284$ vs. .035; $z = 4.16$, $p = 3E-05$), which is consistent with Hypothesis 4b and not 4a (and hence inconsistent with 4c as well). Time 1 numeral comparison was also a stronger unique predictor of mixed-comparison growth than was the reverse (.317 vs. .130; $z = 3.21$, $p = .001$). Time 1 mixed comparison was a stronger unique predictor of dot-comparison growth than was the reverse, albeit nonsignificantly so (.107 vs. .023; $z = 1.37$, $p = .169$).

In sum, symbolic number processing at the beginning of kindergarten was a stronger predictor of growth in nonsymbolic and mixed-format processing than was the other way around. Indeed, symbolic number scores at the beginning of the year were a stronger unique predictor of nonsymbolic and mixed scores at the end of the year than were even nonsymbolic and mixed scores at the beginning of the year, respectively. Results support Hypotheses 3b and 4b and are thus overall more consistent with the SNS \rightarrow AMS view.

Discussion

There is currently considerable discussion about the relative influences of evolutionary and cultural factors in the development of early numerical skills. One part of this debate centers around the relationships between approximate, nonverbal (AMS) and exact, symbolic (SNS) representations of number. Here we examined several hypotheses concerning whether, in the earliest stages of formal education, AMS abilities predict growth in SNS abilities, or the other way around. Moreover, we did so in a manner that takes into account the need to develop an ability to translate between the AMS and the SNS (i.e., mixed comparison). Our data derived from 539 kindergarten children, with AMS, SNS, and mixed-comparison skills assessed at the beginning and end of the academic year. In this way, we provide, to the best of our knowledge, the most comprehensive test to date of the direction of influence between the AMS and the SNS in early formal schooling. For all four hypotheses tested (1–4), results clearly favored the view that SNS abilities at the beginning of kindergarten lay the foundation for improvement in both AMS abilities and the ability to translate between the two systems (results were consistent with Hypotheses 1b–4b and not 1a–4a). Specifically, there was a significant cost of mixing formats present even at the outset of kindergarten, indicating an early dissociation between symbolic and nonsymbolic number systems (Hypothesis 1b). This mixing cost reduced over the course of the school year, indicating an increasing capacity to translate between the two systems (Hypothesis 2b). Growth in this format-mixing was predicted by symbolic but not nonsymbolic ability (Hypothesis 3b). And indeed, SNS ability predicted growth in AMS ability over the course of the year but not the other way around (Hypothesis 4b). More broadly, we conclude that, once one acquires a basic grasp of exact number symbols, it is this exact understanding of number (and repeated practice therewith) that in

fact predicts growth in the AMS. We speculate that one's understanding of approximate, nonverbal quantities is increasingly informed by one's knowledge of and experience with exact symbolic representations of numbers, though the precise mechanism by which this may occur remains unknown. Candidate interpretations and mechanisms are discussed next.

Evidence for the SNS \rightarrow AMS View

To identify the direction of influence between the AMS and the SNS in early education, we examined four hypotheses. First, we examined whether, at the outset of formal education, the AMS and the SNS should be considered separate systems to begin with. We found that, at the beginning of kindergarten (Time 1), scores on the mixed-comparison task were lower than those on either the numeral- or dot-comparison task. This suggests there is an additional cost to translating between the two formats. This result is broadly consistent with those found by Mundy and Gilmore (2009) and Brankaer et al. (2014). Mundy and Gilmore found in their second experiment that children roughly 7–8 years of age struggled with a mapping task (children judged which of two quantities matched the numerosity of a target, where the target was in a different format—symbolic or nonsymbolic—from that of the two options): Average accuracy was 62% (chance = 50%). The same children performed well on standard symbolic–symbolic (90.2% accuracy) and nonsymbolic–nonsymbolic (87.3%) comparison tasks. Because different tasks (matching vs. comparison) were used, it is difficult to directly compare performance on the same-format task with the mixed-format task (as by contrast we were able to do here—note also that a direct comparison between same- and mixed-format tasks was not a stated goal of Mundy & Gilmore, 2009). Nevertheless, it seems that children found the cross-format matching task relatively difficult. More recently, Brankaer et al. found a highly similar result using similar matching and comparison tasks in children approximately 7 and 9 years old.

Taken together with the current results (see especially Figure 2 and tests of Hypotheses 1 and 2 regarding the cost of mixing formats), it seems that the ability to translate between symbolic and nonsymbolic numerical formats is nontrivial, even in children as young as those at the start of kindergarten. This suggests that children's understanding of number symbols may emerge independently of the AMS (for a similar proposal, see, e.g., Carey, 2011; Le Corre & Carey, 2007). At minimum, our results suggest that the SNS is only indirectly linked to the AMS early in development, which is difficult to reconcile with the view that approximate representations of quantity serve as the semantic building blocks of their symbolic counterparts (e.g., Dehaene, 2008; Feigenson et al., 2004, 2013; Piazza, 2010; Piazza et al., 2007; Verguts & Fias, 2004). Furthermore, this result is in direct contrast to the prediction made by Lyons et al. (2012) that the dissociation between symbolic and nonsymbolic numerical representations emerges only later in development with increasing exposure to formal instruction with number symbols.

Indeed, we also found that the cost of translating between the AMS and the SNS *lessened* by the end of kindergarten. This surprising result is in fact broadly consistent with the notion that people's understanding of nonsymbolic quantities is increasingly shaped by (and thus tied to) their understanding of symbolic quantities (and is also consistent with results found by Mix, 2008).

Of course, one could alternatively interpret the reduced mixing cost at Time 2 by positing the reverse: Although perhaps initially distinct, the AMS nevertheless plays a role in shaping the SNS once formal schooling begins. Thus, it is crucial to examine competing longitudinal predictions of growth in the ability to translate between the SNS and the AMS: Is numeral comparison or dot comparison at the beginning of kindergarten a better unique predictor of growth in mixed comparison over the course of the school year?

The longitudinal results shown in Figure 3 clearly support the view that it is the numeral comparison that uniquely predicts growth in mixed comparison and not dot comparison (hence consistent with Hypothesis 3b; see also Figure 3; with partial r s of .317 vs. .023, respectively). Here again, results clearly support the view that early SNS skills predict growth in one's ability to translate between the SNS and the AMS. If children's understanding of AMS-based quantities is increasingly in terms of their grasp of SNS-based quantities, this is exactly what one would expect. We found indirect evidence that, consistent with this view, children were more likely to convert dot arrays to symbolic form than numerals to approximate magnitudes to complete the mixed-comparison task. Systematic underestimation when converting a dot array into a symbolic representation (via noncounting estimation; Izard & Dehaene, 2008) is expected and has been shown in adults to lead to poorer performance in a mixed-comparison task when the dot array is numerically larger than the numeral relative to the opposite (Lyons et al., 2012). This is exactly what we found here: Children were less accurate on mixed-comparison trials where the dot array was numerically greater than the numeral (dots > numeral: 68.9% correct, $SE = 1.2$; numeral > dots: 80.5% correct, $SE = 1.0$; $p < .001$). Even in kindergarten, children seem more inclined to process magnitudes in symbolic form, consistent with the notion that their understanding of the AMS may be increasingly influenced by their grasp of number symbols. It is also worth noting that the ability to map between symbolic and nonsymbolic quantities is a unique predictor of more complex math skills in 6- to 8-year-olds—a result that holds over and above the variance captured by more standard symbolic and nonsymbolic tasks (Mundy & Gilmore, 2009; see also Brankaer et al., 2014). Hence, the ability to map between the AMS and the SNS is both nontrivial and potentially key to the development of math abilities more generally. However, contrary to the currently dominant view, our results suggest that this mapping ability may be facilitated primarily by children's increasing proficiency with numerical symbols and not approximate magnitudes.

The final hypothesis we tested concerns whether AMS skills at the beginning of the year predict growth in SNS skills over the course of the year (AMS \rightarrow SNS; Hypothesis 4a), vice versa (SNS \rightarrow AMS; Hypothesis 4b), or both (AMS \leftrightarrow SNS; Hypothesis 4c). Even if the AMS and the SNS are distinct representational systems, it is possible for the two systems to influence one other. It is thus important to know whether this is the case and, if so, begin to accrue evidence regarding the direction of influence. Again, our results clearly supported the SNS \rightarrow AMS view (Hypothesis 4b). Numeral-comparison scores at the beginning of kindergarten were a stronger unique predictor of growth in dot-comparison scores over the course of the year than was the other way around (from Figure 3; partial r s = .284 vs. .035, respectively). Indeed, numeral comparison (Time 1) was a better unique predictor of dot com-

parison than dot comparison was of itself over the course of the year (from Figure 3; partial r s = .284 vs. .147, respectively). This indicates the SNS provides a stronger influence on the development (at least over the course of kindergarten) of the AMS than does the other way around.

It is important to note that this result is not without precedent. Indeed, this result may in some ways be seen as a replication and extension of two previous studies. In preschoolers, Mussolin et al. (2014) showed that symbolic comparison predicted growth in nonsymbolic comparison but not the other way around. Matejko and Ansari (2016) recently showed a similar result in first graders. It is worth noting that the current sample included roughly an order of magnitude more participants (539 vs. 57 in Mussolin et al., 2014, and 30 in Matejko & Ansari, 2016). Moreover, we also controlled for children's ability to map between the two formats (mixed-comparison performance), and we tested several related hypotheses concerning the ability to map between formats as well. Furthermore, in kindergarteners, Sasanguie, et al. (2014) showed no significant correlation between nonsymbolic comparison performance and symbolic comparison performance 6 months later (the opposite relation could not be assessed, because the authors did not collect symbolic comparison scores at the first time point). Crucially, these studies all clearly converge on the conclusion that it is early symbolic numerical skills (part of the SNS) that facilitate improvement in early nonsymbolic magnitude skills (part of the AMS) and not the other way around. This conclusion is further strengthened by the results testing Hypotheses 1–3 (discussed previously) indicating that one's ability to translate between the SNS and the AMS is both nontrivial and likely facilitated by the fact that approximate magnitudes come to be understood primarily in terms of their symbolic counterparts.

Potential Mechanisms

In the previous section, we reviewed evidence—from both the current study and previous work—that supports the notion that the SNS emerges independently from the AMS and that it is the former that primarily predicts growth in the latter (and not the other way around). Although the evidence for the SNS \rightarrow AMS view outweighs that of the AMS \rightarrow SNS view, the mechanism(s) by which the SNS may influence the AMS remain largely underspecified. In this section we outline two potential mechanisms.

One possibility suggested by Piazza et al. (2013) as well as Mussolin et al. (2016) is that the SNS directly shapes the AMS by improving the representational precision of the latter. Symbols, which tend to be more precise than are their analog counterparts, can help to “sharpen” the nonsymbolic analog magnitude (AMS) representations. These sharper representations then permit more precise discrimination between nonsymbolic quantities, which in turn predicts improved performance on something like a dot-comparison task. On the other hand, Piazza et al. and Mussolin et al. still posited a direct linkage between symbolic and nonsymbolic representations—that is, they are underpinned by the same kind of representation—and indeed the authors suggested a bidirectional influence between the SNS and the AMS. The unidirectionality of longitudinal results here (in particular the lack of evidence supporting the AMS \rightarrow SNS direction; see Figure 3) speak against this assumption, as do our results showing a cost of mixing formats even at the outset of kindergarten (see Figure 2). That said, our

results cannot speak directly to children younger than kindergarteners, for whom some sort of link may have been established.

An alternative perspective is to assume that the SNS and the AMS are completely distinct systems from the outset of development. This would be more consistent with the mixing costs seen here (see Figure 2), the work reviewed earlier by Mundy and Gilmore (2009) and Brankaer et al. (2014), and a growing body of neuroimaging studies showing a distinction between symbolic and nonsymbolic numerical representations, at least in adults (Bulthé et al., 2014, 2015; Damarla et al., 2016; Damarla & Just, 2013; Lyons, Ansari, & Beilock, 2015). According to this view, interrelations among the comparison tasks are due primarily to near-transfer of learning to do the task. The large directional asymmetry is thus simply a function of having more room for growth in the purely symbolic (NC) case. With formal education, children would be expected to improve rapidly on the symbolic comparison task, which can in some ways be seen as having the highest “ceiling” because symbols can be infinitely precise and allow for a rich set of conceptual associations. Indeed, the steepest growth curve in the current data set was seen for the symbolic (NC) task (see Figure 2). According to this view, children with strong symbolic skills are expected to improve most dramatically on the NC task (consistent with the large positive correlation between Time 1 and Time 2 NC), and it is this improvement in learning to perform one type of numerical comparison task that transfers to the others. Unlike the interpretation outlined in the paragraph previous, this view does not require any direct relation between the SNS and the AMS, because mapping between symbolic and analog representations is not required for the SNS to develop. Instead, the SNS develops independently from the AMS; part of this development permits improved symbolic comparison performance, which transfers at the task level to the mixed- and dot-comparison tasks (for a similar suggestion regarding how acquisition of number–word meanings might influence the AMS, see Shusterman et al., 2016).

Castles and Coltheart (2004) made a related argument in the discussion of the literature on the predictors of early reading skills. This argument rests primarily on the notion that a given task does not equal a given process (or representation). Showing that performance on a symbolic number comparison task predicts improvement on a nonsymbolic comparison task does not necessarily imply that SNS representations are fundamentally shaping AMS representations; instead, it may mean that learning how to perform the symbolic task changes how one does the nonsymbolic version of the task. In addressing the surprising finding that reading ability predicts improvement in scores on a phonological awareness task, Castles and Coltheart wrote, “The acquisition of reading skills does not actually change the level or nature of phonological awareness itself. Rather, it influences the way in which children perform phonological awareness tasks” (p. 80). In the current case, learning to process numbers symbolically may change how one does tasks that involve nonsymbolic magnitudes as stimuli. That said, precisely how this change occurs and what it may or may not imply for approximate magnitude representation remains a topic for further research.

Nevertheless, we do not believe our results are merely a fluke of—that is, limited to—number comparison tasks, because related results can be found elsewhere in the numerical literature. For instance, children who did not yet have the cardinality principle were completely at chance on a nonsymbolic comparison task

(Negen & Sarnecka, 2015). The notion here is that only when children have an understanding of the symbolic cardinal label of a set do they understand how to perform a nonsymbolic comparison task. In addition, Mix (2008) tested 3-year-olds’ ability to match sets of objects in terms of their cardinality (is one set of objects numerically equal to another set of objects of a different type). Mix found that children with knowledge of the correct symbolic (verbal) cardinal labels of the sets involved (e.g., “three,” “four”) performed significantly better on a range of nonsymbolic matching tasks. This suggests that children’s ability to compare nonsymbolic quantities is linked to their understanding of the cardinal meaning of number words—which are arguably the first number–symbols most children learn (for a review, see, e.g., Mix et al., 2002). Similarly, Shusterman and colleagues (2016) found longitudinal evidence that children’s acquisition of cardinal understanding of number words preceded an abrupt improvement in performance on AMS comparison acuity (see also Shusterman et al., 2016, for a detailed discussion of the potential mechanisms by which verbal symbols–number words might impact AMS processing). In sum, these results dovetail with those of the current study to suggest that knowledge of and proficiency with the symbolic number system impacts how one processes approximate, analogue magnitudes. Whether acquisition of a symbolic number *system* directly impacts approximate magnitude representation or the broader framework—that is, *system*—by which these magnitudes are (or are not) brought to bear in solving specific tasks remains an interesting avenue for future research.

A point of concern is that processing of nonsymbolic quantities in the subitizing range (1–4) differs from processing of approximate magnitudes outside this range (e.g., Revkin, Piazza, Izard, Cohen, & Dehaene, 2008). In addition, some researchers have suggested that if there is a critical link between symbolic and nonsymbolic representations of quantity, this link is primarily limited to quantities within the subitizing range (e.g., Carey, 2011). This has two implications for the current work. On the one hand, our measure of the AMS may have been “contaminated” by inclusion of trials that use numbers in the subitizing range. In particular, the result that numeral comparison significantly predicted growth in dot comparison may have been inflated by the inclusion of subitizable quantities (a similar argument could be made for the bidirectional relation between numeral and mixed comparisons). On the other hand, the fact that dot comparison did not predict growth in either of the other comparison tasks perhaps speaks against this account. Regardless, we felt it important to test whether the regression results in Figure 3 (and hence our conclusions with respect to Hypotheses 3 and 4) would differ substantially if we limited analyses to only trials comprising stimuli outside the subitizing range (the same limits were imposed on predictors from all three tasks so as not to bias reliability due to the number of trials and ensure that any relations would be exclusive to quantities outside the subitizing range across the board; results are reported in Appendix B). Results show that numeral comparison remained a strong predictor of growth in the other two tasks and that neither dot nor mixed comparison were significant predictors of either one another or numeral comparison. This suggests that the influence of the SNS on the AMS extends beyond the subitizing range, and the asymmetry in the influence between the two systems, may be even stronger for larger numbers. Consistent with the earlier discussion (and with Shusterman et al., 2016), the

relation between the SNS and the AMS is more likely to operate via more general aspects of the respective systems, though further work is certainly needed.

Reconciling Current Results With Previous Work

In the introduction, we indicated that considerable attention has been paid to the view that the AMS plays a crucial role as the foundation for and scaffold of the SNS (Dehaene, 1997, 2008; Feigenson et al., 2004, 2013; Gallistel & Gelman, 2000; Piazza, 2010). Our results are inconsistent with this view; hence, in this section, we briefly examine how our results may potentially be reconciled with previous evidence that has been interpreted in favor of the AMS → SNS view. First, numerous studies have shown that the precision of an individual's AMS is predictive of SNS abilities (for a review and meta-analysis, see Chen & Li, 2014). Nearly all of these studies are cross-sectional in nature, so causal direction cannot be inferred. Moreover, of studies that have examined the relation between AMS and SNS tasks at different time points, few if any include control for the outcome at Time 1 (e.g., Libertus, Feigenson, & Halberda, 2011; Mazzocco, Feigenson, & Halberda, 2011; Wang, Odic, Halberda, & Feigenson, 2016). Inclusion of this control is crucial because it allows one to genuinely predict *growth* in the outcome. For instance, if one uses performance on a dot-comparison task at Time 1 to predict performance on an arithmetic task at Time 2, the direction of influence is still ambiguous. However, if one controls for arithmetic performance at Time 1, then the remaining variance in arithmetic at Time 2 accounted for by dot comparison at Time 1 can be attributed to only what has changed in the arithmetic task between Times 1 and 2; hence, one *can* infer the direction of influence. We included this crucial control here, as did Mussolin et al. (2014) and Matejko and Ansari (2016). In all three cases, the direction of influence was found to be from symbolic to nonsymbolic tasks and not the other way around.

Another line of evidence where causality can be more readily inferred comes from training studies showing that nonsymbolic arithmetic training predicts growth in symbolic arithmetic scores (Hyde et al., 2014; Park & Brannon, 2013, 2014) and that intensive math training does not seem to improve AMS precision (Sullivan, Frank, & Barner, 2016). This certainly suggests that there are instances where nonsymbolic numerical processing can influence SNS processing, though the results mentioned earlier do not appear to extend to training on nonsymbolic comparison (Park & Brannon, 2014). Regardless, it is certainly possible that approximate arithmetic performance at the outset of kindergarten might predict growth in symbolic number skills. Furthermore, it is important to emphasize that our data did not allow us to examine whether dot comparison performance predicts growth in tasks that tap the AMS more directly than do symbolic or mixed comparison, such as nonsymbolic arithmetic.

Finally, it is worth pointing out that in most cases where tasks measuring the AMS (e.g., dot comparison) show a zero-order correlation with symbolic tasks (e.g., symbolic arithmetic), this relation is often completely accounted for by a comparable symbolic task (e.g., numeral comparison; for empirical results, see, e.g., Göbel, Watson, Lervåg, & Hulme, 2014; Lyons, Price, Vaessen, Blomert, & Ansari, 2014; for a recent meta-analysis and review, see, respectively, Schneider et al., 2017, and Merk-

ley & Ansari, 2016). In other words, symbolic numerical tasks tend to be more strongly related to one another than to nonsymbolic numerical tasks. This is broadly consistent with a strong distinction between the SNS and the AMS, which is also consistent with the results here. Moreover, the longitudinal results in the current study suggest that what relation does exist between symbolic and nonsymbolic numerical tasks is reflective primarily of the SNS bearing influence on the AMS and not the other way around.

Conclusion

In sum, we tested several hypotheses concerning the nature and direction of the relation between approximate, nonverbal (AMS) and exact, symbolic (SNS) representations of number. Results converged to show that the two systems are relatively distinct even at the outset of kindergarten. Results also clearly indicated that the SNS unidirectionally predicts growth in the AMS. This asymmetry—in particular the lack of evidence for an AMS → SNS relation—is in direct contrast to the view that the more evolutionarily ancient AMS underpins the culturally acquired SNS. Instead, it appears that culturally acquired number symbols may influence how kindergarteners process nonsymbolic quantities. The precise mechanism by which this process occurs remains unknown. Number symbols may directly change nonsymbolic representations of magnitude, or number symbols may simply afford greater opportunity for improvement in quantity comparison tasks more generally, which transfers to nonsymbolic versions of the task as well. Future work may help elucidate the precise mechanisms. Regardless, the current results may serve to reorient theories about the progression of numerical development especially in the early stages of formal education.

References

- Agrillo, C., & Beran, M. J. (2013). Number without language: Comparative psychology and the evolution of numerical cognition. *Frontiers in Psychology, 4*, 295. <http://dx.doi.org/10.3389/fpsyg.2013.00295>
- Brankaer, C., Ghesquière, P., & De Smedt, B. (2014). Children's mapping between non-symbolic and symbolic numerical magnitudes and its association with timed and untimed tests of mathematics achievement. *PLoS ONE, 9*(4), e93565. <http://dx.doi.org/10.1371/journal.pone.0093565>
- Brannon, E. M., Wusthoff, C. J., Gallistel, C. R., & Gibbon, J. (2001). Numerical subtraction in the pigeon: Evidence for a linear subjective number scale. *Psychological Science, 12*, 238–243. <http://dx.doi.org/10.1111/1467-9280.00342>
- Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *NeuroImage, 87*, 311–322. <http://dx.doi.org/10.1016/j.neuroimage.2013.10.049>
- Bulthé, J., De Smedt, B., & Op de Beeck, H. P. (2015). Visual number beats abstract numerical magnitude: Format-dependent representation of Arabic digits and dot patterns in human parietal cortex. *Journal of Cognitive Neuroscience, 27*, 1376–1387. http://dx.doi.org/10.1162/jocn_a_00787
- Capaldi, E. J., & Miller, D. J. (1988). Counting in rats: Its functional significance and the independent cognitive processes that constitute it. *Journal of Experimental Psychology: Animal Behavior Processes, 14*, 3–17. <http://dx.doi.org/10.1037/0097-7403.14.1.3>
- Carey, S. (2011). *The origin of concepts*. <http://dx.doi.org/10.1017/S0140525X10000919>

- Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition*, *91*, 77–111. [http://dx.doi.org/10.1016/S0010-0277\(03\)00164-1](http://dx.doi.org/10.1016/S0010-0277(03)00164-1)
- Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, *148*, 163–172. <http://dx.doi.org/10.1016/j.actpsy.2014.01.016>
- Crollen, V., Castronovo, J., & Seron, X. (2011). Under- and over-estimation: A bi-directional mapping process between symbolic and non-symbolic representations of number? *Experimental Psychology*, *58*, 39–49. <http://dx.doi.org/10.1027/1618-3169/a000064>
- Damarla, S. R., Cherkassky, V. L., & Just, M. A. (2016). Modality-independent representations of small quantities based on brain activation patterns. *Human Brain Mapping*, *37*, 1296–1307. <http://dx.doi.org/10.1002/hbm.23102>
- Damarla, S. R., & Just, M. A. (2013). Decoding the representation of numerical values from brain activation patterns. *Human Brain Mapping*, *34*, 2624–2634. <http://dx.doi.org/10.1002/hbm.22087>
- Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. <http://dx.doi.org/10.1017/S0790966700006248>
- Dehaene, S. (2008). Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation. In P. Haggard & Y. Rossetti (Eds.), *Attention and Performance XXII: Sensorimotor foundations of higher cognition* (pp. 527–574). New York, NY: Oxford University Press.
- Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, *56*, 384–398. <http://dx.doi.org/10.1016/j.neuron.2007.10.004>
- 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>
- Gallistel, C. R., & Gelman, I. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, *4*, 59–65. [http://dx.doi.org/10.1016/S1364-6613\(99\)01424-2](http://dx.doi.org/10.1016/S1364-6613(99)01424-2)
- 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>
- 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>
- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*, 102–116. <http://dx.doi.org/10.1037/a0038889>
- Hyde, D. C., Khanum, S., & Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition*, *131*, 92–107. <http://dx.doi.org/10.1016/j.cognition.2013.12.007>
- Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*, 1221–1247. <http://dx.doi.org/10.1016/j.cognition.2007.06.004>
- 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. <http://dx.doi.org/10.1016/j.cognition.2006.10.005>
- Leibovich, T., & Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Frontiers in Psychology*, *4*, 375. <http://dx.doi.org/10.3389/fpsyg.2013.00375>
- 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. (2003). Origins of number sense: Large-number discrimination in human infants. *Psychological Science*, *14*, 396–401. <http://dx.doi.org/10.1111/1467-9280.01453>
- Lyons, I. M., & Ansari, D. (2015). Foundations of children's numerical and mathematical skills: The roles of symbolic and nonsymbolic representations of numerical magnitude. *Advances in Child Development and Behavior*, *48*, 93–116. <http://dx.doi.org/10.1016/bs.acdb.2014.11.003>
- 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., Ansari, D., & Beilock, S. L. (2015). Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain. *Human Brain Mapping*, *36*, 475–488. <http://dx.doi.org/10.1002/hbm.22641>
- Lyons, I. M., Nuerk, H. C., & Ansari, D. (2015). Rethinking the implications of numerical ratio effects for understanding the development of representational precision and numerical processing across formats. *Journal of Experimental Psychology: General*, *144*, 1021–1035. <http://dx.doi.org/10.1037/xge0000094>
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in Grades 1–6. *Developmental Science*, *17*, 714–726. <http://dx.doi.org/10.1111/desc.12152>
- Matejko, A. A., & Ansari, D. (2016). Trajectories of symbolic and non-symbolic magnitude processing in the first year of formal schooling. *PLoS ONE*, *11*(3), e0149863. <http://dx.doi.org/10.1371/journal.pone.0149863>
- Matthews, P. G., Lewis, M. R., & Hubbard, E. M. (2016). Individual differences in nonsymbolic ratio processing predict symbolic math performance. *Psychological Science*, *27*, 191–202. <http://dx.doi.org/10.1177/0956797615617799>
- Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, *6*(9), e23749. <http://dx.doi.org/10.1371/journal.pone.0023749>
- McCrink, K., & Spelke, E. S. (2010). Core multiplication in childhood. *Cognition*, *116*, 204–216. <http://dx.doi.org/10.1016/j.cognition.2010.05.003>
- McCrink, K., & Spelke, E. S. (2016). Non-symbolic division in childhood. *Journal of Experimental Child Psychology*, *142*, 66–82. <http://dx.doi.org/10.1016/j.jecp.2015.09.015>
- Merkley, R., & Ansari, D. (2016). Why numerical symbols count in the development of mathematical skills: Evidence from brain and behavior. *Current Opinion in Behavioral Sciences*, *10*, 14–20. <http://dx.doi.org/10.1016/j.cobeha.2016.04.006>
- Mix, K. S. (2008). Surface similarity and label knowledge impact early numerical comparisons. *British Journal of Developmental Psychology*, *26*, 13–32. <http://dx.doi.org/10.1348/026151007X189109>
- Mix, K. S., Huttenlocher, J., & Levine, S. C. (1996). Do preschool children recognize auditory-visual numerical correspondences? *Child Development*, *67*, 1592–1608. <http://dx.doi.org/10.2307/1131720>
- Mix, K. S., Huttenlocher, J., & Levine, S. C. (2002). *Quantitative development in infancy and early childhood*. <http://dx.doi.org/10.1093/acprof:oso/9780195123005.001.0001>
- 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>
- Murphy, K. R., & Davidshofer, C. O. (2004). *Psychological testing: Principles and applications* (6th ed.). London, United Kingdom: Pearson.
- Mussolin, C., Nys, J., Content, A., & Leybaert, J. (2014). Symbolic number abilities predict later approximate number system acuity in preschool

- children. *PLoS ONE*, 9(3), e91839. <http://dx.doi.org/10.1371/journal.pone.0091839>
- 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>
- Negen, J., & Sarnecka, B. W. (2015). Is there really a link between exact-number knowledge and approximate number system acuity in young children? *British Journal of Developmental Psychology*, 33, 92–105. <http://dx.doi.org/10.1111/bjdp.12071>
- Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, 32, 185–208. <http://dx.doi.org/10.1146/annurev.neuro.051508.135550>
- Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLoS ONE*, 8(7), e67918. <http://dx.doi.org/10.1371/journal.pone.0067918>
- Oppen-Rhein, R., & Strimmer, K. (2007). Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach. *Statistical Applications in Genetics and Molecular Biology*, 6. Advance online publication. <http://dx.doi.org/10.2202/1544-6115.1252>
- Pahl, M., Si, A., & Zhang, S. (2013). Numerical cognition in bees and other insects. *Frontiers in Psychology*, 4, 162. <http://dx.doi.org/10.3389/fpsyg.2013.00162>
- Park, J., & Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychological Science*, 24, 2013–2019. <http://dx.doi.org/10.1177/0956797613482944>
- Park, J., & Brannon, E. M. (2014). Improving arithmetic performance with number sense training: An investigation of underlying mechanism. *Cognition*, 133, 188–200. <http://dx.doi.org/10.1016/j.cognition.2014.06.011>
- Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences*, 14, 542–551. <http://dx.doi.org/10.1016/j.tics.2010.09.008>
- Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychological Science*, 24, 1037–1043. <http://dx.doi.org/10.1177/0956797612464057>
- Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, 53, 293–305. <http://dx.doi.org/10.1016/j.neuron.2006.11.022>
- Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science*, 19, 607–614. <http://dx.doi.org/10.1111/j.1467-9280.2008.02130.x>
- Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin*, 88, 245–258. <http://dx.doi.org/10.1037/0033-2909.88.2.245>
- Rowley, G. L., & Traub, R. E. (1977). Formula scoring, number-right scoring, and test-taking strategy. *Journal of Educational Measurement*, 14, 15–22. <http://dx.doi.org/10.1111/j.1745-3984.1977.tb00024.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, 271–280. <http://dx.doi.org/10.1080/17470218.2013.803581>
- Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, 20, e12372. <http://dx.doi.org/10.1111/desc.12372>
- Shusterman, A., Slusser, E., Halberda, J., & Odic, D. (2016). Acquisition of the cardinal principle coincides with improvement in approximate number system acuity in preschoolers. *PLoS ONE*, 11(4), e0153072. <http://dx.doi.org/10.1371/journal.pone.0153072>
- Sullivan, J., Frank, M. C., & Barner, D. (2016). Intensive Math Training Does not Affect Approximate Number Acuity: Evidence From a Three-Year Longitudinal Curriculum Intervention. *Journal of Numerical Cognition*, 2, 57–76. <http://dx.doi.org/10.5964/jnc.v2i2.19>
- Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of Cognitive Neuroscience*, 16, 1493–1504. <http://dx.doi.org/10.1162/0898929042568497>
- Wang, J. J., Odic, D., Halberda, J., & Feigenson, L. (2016). Changing the precision of preschoolers' approximate number system representations changes their symbolic math performance. *Journal of Experimental Child Psychology*, 147, 82–99. <http://dx.doi.org/10.1016/j.jecp.2016.03.002>
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74(1), B1–B11. [http://dx.doi.org/10.1016/S0010-0277\(99\)00066-9](http://dx.doi.org/10.1016/S0010-0277(99)00066-9)
- Zhang, J., & Norman, D. A. (1995). A representational analysis of numeration systems. *Cognition*, 57, 271–295. [http://dx.doi.org/10.1016/0010-0277\(95\)00674-3](http://dx.doi.org/10.1016/0010-0277(95)00674-3)

(Appendices follow)

Appendix A

Full Longitudinal Regression Results

Appendix A gives full regression results for the three regression models used to test the longitudinal hypotheses. Note that these models were used to produce the partial r values (and corresponding p values) depicted in Figure 3.

Table A1
Outcome: Numeral-Comparison Task at Time 2

Predictor	b	SE	t	p	r_o	r_p
NC1	.544	.048	11.377	6E-27	.716	.444
DC1	.063	.077	.813	.416	.548	.035
MC1	.232	.077	3.005	.003	.623	.130
Sex	-1.357	.878	-1.546	.123	-.050	-.067
Absent %	2.521	6.166	.409	.683	-.053	.018
Test interval	.036	.032	1.121	.263	.034	.049
Born in Canada	1.478	1.346	1.099	.272	.014	.048
School SES	.021	.481	.045	.964	.216	.002
NC1 > chance	-2.824	1.860	-1.519	.129	.303	-.066
DC1 > chance	-.627	2.430	-.258	.796	.170	-.011
MC1 > chance	.910	1.749	.520	.603	.325	.023
NC2 > chance	21.731	3.242	6.702	5E-11	.318	.281
Constant	-8.928	6.857	-1.302	.193		
Adjusted R^2	.564					
Residual df	526					

Note. NC = numeral comparison; DC = dot comparison; MC = mixed comparison; 1 = Time 1 (fall); 2 = Time 2 (spring); SES = socioeconomic status.

Table A2
Outcome: Dot-Comparison Task at Time 2

Predictor	b	SE	t	p	r_o	r_p
NC1	.257	.038	6.805	3E-11	.590	.284
DC1	.209	.061	3.413	7E-04	.533	.147
MC1	.151	.061	2.475	.014	.526	.107
Sex	.816	.691	1.181	.238	.025	.051
Absent %	-.181	4.867	-.037	.970	-.042	-.002
Test interval	-.005	.025	-.215	.830	-.029	-.009
Born in Canada	.744	1.062	.700	.484	.027	.031
School SES	-.310	.380	-.814	.416	.166	-.035
NC1 > chance	-1.651	1.445	-1.142	.254	.210	-.050
DC1 > chance	-.570	1.905	-.299	.765	.180	-.013
MC1 > chance	-2.169	1.377	-1.575	.116	.185	-.069
NC2 > chance	19.671	2.865	6.867	2E-11	.303	.287
Constant	-1.898	5.832	-.325	.745		
Adjusted R^2	.430					
Residual df	526					

Note. NC = numeral comparison; DC = dot comparison; MC = mixed comparison; 1 = Time 1 (fall); 2 = Time 2 (spring); SES = socioeconomic status.

(Appendices continue)

Table A3
Outcome: Mixed-Comparison Task at Time 2

Predictor	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>r</i> ₀	<i>r</i> _p
NC1	.303	.039	7.675	8E-14	.624	.317
DC1	.034	.064	.525	.600	.512	.023
MC1	.295	.064	4.636	4E-06	.595	.198
Sex	-1.321	.722	-1.829	.068	-.069	-.079
Absent %	-2.897	5.090	-.569	.570	-.061	-.025
Test interval	-.024	.026	-.910	.363	-.047	-.040
Born in Canada	1.555	1.110	1.401	.162	.052	.061
School SES	-.298	.397	-.751	.453	.173	-.033
NC1 > chance	-2.860	1.514	-1.889	.059	.223	-.082
DC1 > chance	1.144	1.993	.574	.566	.180	.025
MC1 > chance	-1.563	1.444	-1.082	.280	.272	-.047
NC2 > chance	16.363	2.246	7.284	1E-12	.349	.303
Constant	3.783	5.596	.676	.499		
Adjusted <i>R</i> ²	.479					
Residual <i>df</i>	526					

Note. NC = numeral comparison; DC = dot comparison; MC = mixed comparison; 1 = Time 1 (fall); 2 = Time 2 (spring); SES = socioeconomic status.

Appendix B

Full Regression Results (Large Items Only)

Here we test whether inclusion of trials involving smaller quantities may have inflated the predictive capacity of the mixed and numeral comparison tasks and may have deflated the predictive capacity of the dot comparison task. We tested this by including only trials where all quantities were outside the subitizing range

(>5) for all three critical predictors (the three comparison tasks at Time 1). Results in the tables show that they did not. Tables are organized in the same manner as in Appendix A for easy comparison. Slight variations in degrees of freedom are due to a few children not having attempted any qualifying trials.

Table B1
Outcome: Numeral-Comparison Task at Time 2

Predictor	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>r</i> ₀	<i>r</i> _p
NC1	1.494	.099	15.134	3E-43	.675	.551
DC1	.050	.171	.295	.768	.236	.013
MC1	.224	.157	1.426	.154	.412	.062
Sex	-.680	.932	-.730	.466	-.051	-.032
Absent %	-2.765	6.533	-.423	.672	-.052	-.018
Test interval	.010	.034	.309	.758	.034	.013
Born in Canada	.928	1.412	.657	.512	.014	.029
School SES	.667	.500	1.334	.183	.217	.058
NC1 > chance	.699	1.901	.368	.713	.304	.016
DC1 > chance	2.102	2.409	.872	.383	.170	.038
MC1 > chance	3.970	1.655	2.399	.017	.325	.104
NC2 > chance	24.323	3.401	7.151	3E-12	.318	.298
Constant	21.216	6.662	3.185	.002		
Adjusted <i>R</i> ²	.519					
Residual <i>df</i>	524					

Note. NC = numeral comparison; DC = dot comparison; MC = mixed comparison; 1 = Time 1 (fall); 2 = Time 2 (spring); SES = socioeconomic status.

(Appendices continue)

Table B2

Outcome: Dot-Comparison Task at Time 2

Predictor	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>r</i> _o	<i>r</i> _p
NC1	.824	.077	10.657	4E-24	.548	.422
DC1	.329	.134	2.461	.014	.268	.107
MC1	.172	.123	1.395	.164	.353	.061
Sex	1.413	.728	1.942	.053	.028	.085
Absent %	-1.752	5.116	-.342	.732	-.036	-.015
Test interval	-.024	.026	-.899	.369	-.029	-.039
Born in Canada	.161	1.106	.146	.884	.027	.006
School SES	.201	.393	.512	.609	.168	.022
NC1 > chance	.625	1.461	.428	.669	.211	.019
DC1 > chance	2.161	1.874	1.153	.249	.181	.050
MC1 > chance	.180	1.285	.140	.889	.186	.006
NC2 > chance	22.353	2.978	7.505	3E-13	.304	.312
Constant	21.319	5.210	4.092	5E-05		
Adjusted <i>R</i> ²	.379					
Residual <i>df</i>	524					

Note. NC = numeral comparison; DC = dot comparison; MC = mixed comparison; 1 = Time 1 (fall); 2 = Time 2 (spring); SES = socioeconomic status.

Table B3

Outcome: Mixed-Comparison Task at Time 2

Predictor	<i>b</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>r</i> _o	<i>r</i> _p
NC1	.908	.082	11.140	5E-26	.580	.438
DC1	.074	.141	.526	.599	.233	.023
MC1	.354	.130	2.728	.007	.415	.118
Sex	-.899	.767	-1.172	.242	-.069	-.051
Absent %	-6.336	5.392	-1.175	.241	-.061	-.051
Test interval	-.046	.028	-1.648	.100	-.046	-.072
Born in Canada	1.183	1.165	1.015	.310	.052	.044
School SES	.118	.413	.286	.775	.173	.013
NC1 > chance	-.473	1.545	-.306	.760	.223	-.013
DC1 > chance	2.656	1.973	1.346	.179	.180	.059
MC1 > chance	1.573	1.364	1.153	.249	.272	.050
NC2 > chance	18.582	2.351	7.902	2E-14	.349	.326
Constant	25.652	5.489	4.673	4E-06		
Adjusted <i>R</i> ²	.426					
Residual <i>df</i>	524					

Note. NC = numeral comparison; DC = dot comparison; MC = mixed comparison; 1 = Time 1 (fall); 2 = Time 2 (spring); SES = socioeconomic status.

(Appendices continue)

Appendix C

Correcting Longitudinal Regression Results for Variation in Task Reliability

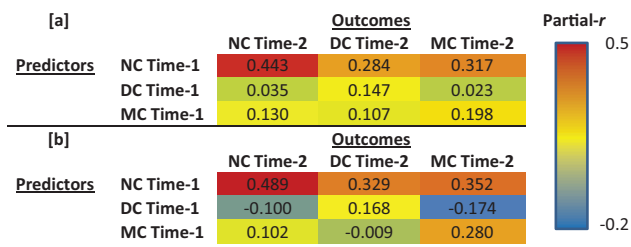


Figure C1. Section [a] gives the partial correlations depicted in Figure 3. Section [b] gives these correlations computed only after the original relations between variables were first disattenuated based on relative reliabilities. The central results reported in the main text remain largely unchanged. See the online article for the color version of this figure.

Here we report the critical partial correlations shown in Figure 3 but recomputed after disattenuating relations between the critical variables with respect to their relative reliabilities (see Figure C1). In the main text, we ran three regression analyses—one with each of the three comparison tasks at Time 2 as the dependent variable. The three comparison tasks at Time 1 were always the critical independent variables. We computed the partial correlation matrix between the four critical variables—for example, Time 1 numeral comparison, Time 1 dot comparison, Time 1 mixed comparison,

Time 2 numeral comparison—after removing the influence of all covariates. Note that at this stage, the correlations between critical variables were residualized only with respect to the covariates and not yet with respect to one another. These correlation matrices were then disattenuated: $r'(xy) = r(xy)/\sqrt{(xr \times yr)}$, where $r(xy)$ is the correlation between variables x and y , and xr and yr are the reliability estimates for x and y , respectively (e.g., Murphy & Davidshofer, 2004). Reliability values were taken from the main text (see the Reliability section). Because reliability was less than perfect (<1) for all tasks, this meant that correlations would be expected to increase in all cases; crucially, however, those involving variables with lower reliability (e.g., between DC and MC) would be expected to increase more so than those involving higher reliability. To residualize the critical variables with respect to one another (i.e., to compute unique contributions of each task at Time 1 to growth in the dependent variable (as is depicted in Figure 3), we “reduced” the correlation using a pseudo-inverse procedure that removes all mutual influence (resulting in a matrix of partial correlations akin to what one would get from a multiple regression (Opgen-Rhein & Strimmer, 2007).

Received September 23, 2016

Revision received August 7, 2017

Accepted August 23, 2017 ■