**Children’s arithmetic development: it is number knowledge, not the approximate number sense, that counts**

Silke M. Göbel1, Sarah E. Watson1, Arne Lervåg2 & Charles Hulme3

1. University of York, UK
2. University of Oslo, Norway
3. University College London, UK

Corresponding Author:

Charles Hulme

Division of Psychology and Language Sciences

2 Wakefield Street

London

WC1N 1PF

Email: c.hulme@ucl.ac.uk

**Acknowledgements:** Sarah Watson was supported by a BBSRC studentship.

Word Count: 3988 (main text)

**Abstract**

The authors present the results of an 11-month longitudinal study (beginning when children were 6 years old) focussing on measures of the approximate number sense (ANS) and knowledge of the Arabic numeral system as possible influences on the development of arithmetic skills. Multiple measures of symbolic and non-symbolic magnitude judgement were shown to define a unitary factor that appears to index the efficiency of an ANS system and which is a strong longitudinal correlate of arithmetic skills. However, longitudinal path models showed that knowledge of Arabic numerals at 6 years was a powerful longitudinal predictor of the growth in arithmetic skills while variations in magnitude comparison ability played no additional role in predicting variations in arithmetic skills. These results suggest that verbal processes concerned with learning the labels for Arabic numerals, and the ability to translate between Arabic numerals and verbal codes, place critical constraints on arithmetic development. [148 words]

Keywords: cognitive development, numerical development, arithmetic, approximate number sense, symbolic number processing

Children’s arithmetic development: it is number knowledge, not the approximate number sense, that counts

The mastery of basic arithmetic skills is a key goal of early education, but so far our understanding of the cognitive factors underlying arithmetic development is arguably limited. This paper reports a large-scale longitudinal study which explores the cognitive bases of early arithmetic development.

One dominant theory is that the development of arithmetic skills depends upon the integrity of an innate approximate number sense (ANS) (Dehaene, 1992). The precision of coding within the ANS is typically assessed by the accuracy and speed of discriminating the numerosity of groups of objects (Barth, Kanwisher & Spelke, 2003; Piazza et al., 2010). Such discriminations can be made more quickly than is possible by counting and vary in difficulty in proportion to the ratio between the numerosities. These findings provide evidence for the operation of an “approximate” number sense that is not based on counting. The precision of ANS coding increases with age: six-month-old infants can distinguish 8 from 16 objects but not 8 from 12 objects (Xu & Spelke, 2000), while most adults can distinguish 9 from 10 objects without counting (Halberda & Feigenson, 2008).

It has been suggested that the ANS provides the foundation for the later development of abstract, symbolic number concepts which in turn underpin the development of arithmetic skills (e.g. Piazza & Dehaene, 2004). If this is the case variations in the precision of coding within the ANS should be related to individual differences in arithmetic skills in children. However, De Smedt, Noël, Gilmore & Ansari (2013) reviewed existing studies and concluded that evidence for such a relationship was lacking (7 studies with children found such a relationship while 11 did not). Furthermore, the majority of studies assess concurrent correlations which may reflect the fact that learning arithmetic leads to improvements in the precision of coding in the ANS (Halberda, Mazzocco, & Feigenson, 2008). For example, Piazza, Pica, Izard, Spelke and Dehaene (2013) found that the precision of coding in the ANS was greater in participants who had received formal education compared with those who had not.

Crucially, we need evidence from longitudinal studies that assess whether variations in the precision of coding in the ANS are predictive of later arithmetic skills. The few longitudinal studies to date have generated mixed results. Mazzocco, Feigenson & Halberda (2011) found a strong relationship between accuracy on a numerosity discrimination task (at age 4 years 2 months) and performance on a standardized measure of arithmetic some 2.5 years later. However, the sample here was very small (N=17) and numerosity discrimination was the only measure taken at preschool, so it is impossible to assess the specificity of this relationship.

Libertus, Feigenson, and Halberda (2013) reported that accuracy (*r* =.44) and reaction time (*r* = -.28) on a numerosity discrimination task (at 4 years 2 months) correlated with scores on an arithmetic test 8 months later. However, using directly comparable measures in the same age group Fuhs and McNeil (2013) found a nonsignificant concurrent correlation between numerosity discrimination and arithmetic scores (*r* = .19). Two other longitudinal studies, with slightly older children, found weak relationships between the ANS and arithmetic skills. Desoete, Ceulemans, De Werdt, and Pieters (2012) found a weak but significant correlation (*r* = .16; N = 395) between non-symbolic numerosity discrimination in kindergarten and calculation scores one year later, while Sasanguie, Göbel, Smets & Reynvoet (2012) found a correlation of similar magnitude (*r* = 0.17).

Evidence for the ANS playing a role in early arithmetic development is clearly mixed, and a major aim of the current study was to clarify its role. Whatever the role of the ANS in arithmetic development, it is also clear that learning the symbol set (the Arabic numeral system) is likely to be a major influence on early arithmetic development (e.g. Krajewski & Schneider, 2009; Purpura, Baroody, & Lonigan, 2013). Accordingly, we will assess children’s knowledge of Arabic numerals (the ability to match numerals to their spoken form) as well as their ability to make speeded number comparisons with Arabic digits. Individual differences in judging the magnitudes represented by Arabic digits (symbolic number comparison skills) are concurrent correlates of arithmetic, but not reading skills (Bugden & Ansari, 2011; Holloway & Ansari, 2009; Sansanguie et al., 2012). Several studies have also shown significant longitudinal relationships between symbolic magnitude comparison assessed in kindergarden or grade 1 and later arithmetic skills (see DeSmedt et al., 2013, for a review). A critical question which we address here is how non-symbolic (numerosity) and symbolic (digit magnitude) judgement tasks relate to each other, and how they function as longitudinal predictors of arithmetic development. By using multiple measures of both non-symbolic (dots) and symbolic (Arabic numeral) numerosity discrimination we can assess the underlying structure of the ANS and how it relates longitudinally to the growth of arithmetic skills. In addition to assessing judgements of the comparative magnitude of digit pairs (which is bigger: 3 or 5?) we use a directly analogous task requiring judgements of the order of letters in the alphabet (which letter comes later in the alphabet: c or e?). This task requires access to an ordered representation of symbols that is directly analogous to the digit judgement task, but does not require access to numerical or magnitude information.

The aim of the current study was to identify the longitudinal predictors of growth in arithmetic during a period (6 years 3 months to 7 years 2 moths) where there is rapid development in basic arithmetic skills. Using multiple measures of magnitude comparison, combined with latent variable analyses, allows us to assess the underlying factor structure of the constructs assessed and to maximize the reliability of the measurements. Our first major aim was to assess the factorial structure and reliability of measures of a hypothetical ANS system (do different measures of this construct cohere to define a unitary construct and can it be measured reliably?). Our second aim was to identify possible predictors of growth in arithmetic and possible reciprocal relationships between the development of arithmetic and a hypothetical ANS system.

**Method**

**Participants**

All children in Year 1 in four UK primary schools (eight classes) were invited to take part; 173 children (mean age 6 years 3 months; 97 boys and 76 girls) participated at time 1 and 165 (93 boys and 72 girls) were retested at time 2 approximately 11 months later.

**Materials and Procedure**

All tests were administered as paper and pencil measures to whole class groups in a fixed order in sessions of 1 hour duration (3 sessions at time 1; 2 sessions at time 2). The measures reported here form part of a larger test battery.

At time 1 nonverbal ability, vocabulary knowledge, number identification, letter comparison, magnitude comparison and arithmetic were assessed (and the latter two skills were re-assessed at Time 2)

***Nonverbal ability.*** To assess general cognitive ability, sets A, B, and C of the Raven’s Standard Progressive Matrices – Plus (Raven’s SPM-Plus; Raven, Raven, & Court, 1998) were administered. Children were given an incomplete matrix and asked to choose the form that completed the matrix. The first two items were practice items. One point was given per correct response (maximum score = 34).

***Vocabulary knowledge.*** We administered 36 items taken from sets five to seven of the British Picture Vocabulary Scale 3rd Edition (BPVS III; Dunn, Dunn & Styles, 2009; with publisher’s permission). The pictures were shown to the class using an electronic whiteboard. Children indicated in their printed answer booklet which of four pictures best matched the spoken target word. One point was given for each item correct (maximum score = 36).

***Number identification.*** A number identification task (Number ID) with eight items was constructed to assess children’s ability to identify one, two, and three-digit Arabic numerals. The experimenter said the target number aloud (e.g. ‘fourteen’) and the children attempted to identify the corresponding Arabic numeral out of four or five presented response options on the answer sheet. The first item was a single-digit number (6) followed by four two-digit numbers (14, 28, 52, and 76) and three three-digit numbers (163, 235, and 427). Distractors were chosen based on visual similarity to the target number and on common errors with place-value (e.g. target number ‘163’, choices: 13, 10063, 136, 16, 163). One point was given per correct item (maximum score = 8).

**Comparison tasks.** All comparison tasks were presented in an A5 answer booklet. Six pairs of items were presented per page with one pair on each row. Each comparison task started with an instruction page with an example. Children were told that they had to answer as many items as quickly as possible by ticking the larger item in a pair without leaving out any pairs until the experimenter said “stop”. Children were only allowed to turn over to the test pages once the experimenter said “go”. Children were stopped after 30s and the next comparison task was presented. One point was given per correct response. These measures therefore yield an estimate of efficiency (the number of correct responses per unit time).

***Magnitude comparison.***The *Digit comparison* tasks used the digits 1 to 9 (Calibri, font size 48). Children were presented with two versions of the digit comparison task: close and far. Pairs of digits in the close version (Digits close) had a numerical distance of either 1 or 2; in the far version (Digits far) the numerical distance between the digits in each pair was 5, 6 or 7. Each pair was matched on the total problem size of the two digits in the close and far versions of the task, for example the pair (5:3) in the close version was replaced by the pair (7:1) in the far version.

The *Non-symbolic number comparison* tasks involved arrays of between 5 and 40 black squares arranged randomly within a 2.5 cm2 box on a white background. There were two versions of this task (same size and same surface area).

In the same size task squares of the same size were used for each item pair and the number of squares presented in each box ranged from 5 to 13. The close comparison version (Non-symbolic Close Same Size (NSC-SS)) involved comparing two numerosities differing by 1 or 2, while the far comparison version (Non-symbolic Far Same Size (NSF-SS)) involved comparing numerosities differing by 5, 6 or 7. The two versions were matched on the total problem size of the two numerosities being compared, for example the close comparison item (10:8) was matched to the far comparison item (12:6).

In the same surface area task the number of squares presented ranged from 20 to 40 but in this case the total surface area of the squares was matched in each item pair (to prevent discriminations being based on surface area rather than numerosity). The ratios between the numerosities were 3:4, 5:6, and 7:8 and each of these tasks was presented separately (3:4 SA, 5:6 SA, 7:8 SA). The numbers 20 through 30 were used as a baseline and the nearest whole number at each ratio was calculated. For example 20 was compared to 27 (3:4), 24 (5:6), and 23 (7:8).

***Letter comparison.*** The letter comparison task consisted of pairs of letters ranging from a to i (Comic Sans MS, font size 48, lower case). Children had to indicate which of the two letters came later in the alphabet by ticking the correct letter. Problems in the letter comparison task were matched to those in the digit comparison task (digit 1 was replaced with letter a, digit 2 with letter b etc.). There were two versions of the letter comparison task: Letter pairs in the letters close version were either adjacent to each other in the alphabet (e.g. a b) or were just one letter apart (e.g. a c), in the letters far version the distance in position between the letters was 5, 6 or 7 (e.g. a g).

***Arithmetic.*** Arithmetic was assessed using the Numerical Operations subtest of the Wechsler Individual Achievement Test-Second UK Edition (WIAT-IIUK; Wechsler, 2005) adapted for group use. This subtest begins with six items that simply involve identifying and writing Arabic numerals, and counting dots. Because we wanted a measure of conventional arithmetic skill, for the present analyses responses to these first six items were excluded. At time 1 there were 9 items that involved conventional arithmetical calculation (addition, subtraction, and multiplication problems). At time 2 the same items were re-administered together with 3 more difficult items (maximum score 9 at time 1 and 12 at time 2). Children were guided through the first six items but were restricted to 15 minutes to complete the conventional arithmetical calculation problems.

**Results**

The means, standard deviations and reliabilities for all variables at both time points are shown in Table 1.

A number of features of the data from the digit and non-symbolic magnitude comparison tasks are worthy of comment. For the digit and dot array comparison tasks (NSF-SS; NSC-SS; 3:4 SA; 5:6 SA; 7:8 SA) there are clear “distance effects” with children comparing pairs of items that are numerically further apart more efficiently than items that are numerically more similar. (All pairwise comparisons between means in each task (NSF-SS vs NSC-SS; 3:4 SA vs. 5:6 SA; 5:6 SA vs. 7:8 SA) showed highly reliable differences at both time 1 and time 2). This pattern (better performance on numerically more distant pairs together with rapid performance (average solution time of 1.36s per item in the NSF-SS task at time 1)) provides evidence that the non-symbolic comparison tasks are indeed tapping an ANS system and that children are not performing these tasks by counting.

Structural equation models of these data were estimated with Mplus 7 (Muthén & Muthén, 1998-2013) with missing values being handled with Robust Full Information Maximum Likelihood estimation.

Figure 1 shows a latent variable path model in which variations in arithmetic at Time 2 were predicted from all constructs measured at Time 1 (arithmetic, magnitude comparison, letter comparison, number identification, vocabulary skills, nonverbal abilities and age). In this model, all seven non-symbolic and symbolic numerical magnitude judgment tasks load on a single latent magnitude comparison factor while the two letter judgment tasks load on a different factor. (Preliminary analyses showed that this two factor model for all comparison tasks fitted the data significantly better than a model equivalent to a one factor solution where the correlation between the magnitude comparison and letter comparison factors was fixed to one, S-B *χ2*Δ(1) = 13.40, *p* = .001. Furthermore, when considering the symbolic(digit) and non-symbolic magnitude comparison tasks alone, there was no difference between the one-factor magnitude model and a nested model where the symbolic and non-symbolic comparison tasks loaded on one factor each, S-B *χ2*Δ(1) = 0.073, *p* = .787). Because arithmetic, number identification, nonverbal ability and vocabulary were each assessed by only one indicator (WIAT, Number ID, Raven and BPVS), to avoid distortions caused by measurement error, we pre-specified the error variance for these measures based on their reliabilities. The model fitted the data very well, *χ2* (68) = 95.12, *p* = .017, Root Mean Square Error of Approximation (RSMEA) = .048 (90% CI = .021-.070), Comparative Fit Index (CFI) = .971, Standardized Root Mean Residuals (SRMR) = .040, which confirms that the factor structure specified in the measurement model for the Time 1 measures is satisfactory. This model shows that there were two unique predictors of individual differences in arithmetic skill at Time 2: arithmetic at Time 1 and number identification. Overall the model explained 86% of the variance in arithmetic.

The pattern of correlations between the latent variables in the model is highly informative (see Table 2). First, our measure of the ANS (magnitude comparison) at time 1 is a strong longitudinal correlate of arithmetic skill at time 2. In fact the bivariate longitudinal correlation between magnitude comparison and arithmetic skills here is stronger (*r* = .60 )than in previous longitudinal studies (where correlations range between .16 and .44 (see the description of these studies on p.4)). The strength of this correlation likely reflects the fact that it is based on the true score variance in these constructs. However, the equivalent measure involving letter comparison at time 1 has an identical longitudinal correlation with arithmetic at time 2. Although the magnitude and letter comparison tasks have identical task requirements the letter comparison task clearly does not require access to numerical magnitudes and as the analyses presented earlier confirm, the letter and magnitude judgement measures define separable factors. This finding therefore clearly questions whether magnitude comparison plays any specific role in explaining individual differences in arithmetic skill in this age range. Finally, the strongest longitudinal correlate of arithmetic at time 2 is the number identification task. In summary, these correlations, and the model in Figure 1, show that an ANS construct can be defined and measured with high reliability but that it does not predict unique variance in the growth of arithmetic skills after other related measures are taken into consideration.

The correlations in Table 2 show a high degree of shared variance between arithmetic, letter comparison, magnitude comparison and number identification at Time 1 as predictors of arithmetic at time 2. To address this issue, and explore the relative contributions of these variables in more detail, we used Cholesky factoring (equivalent to hierarchical regression with latent variables) to estimate the unique variance in arithmetic at Time 2 accounted for by the three theoretically important constructs (magnitude comparison, letter comparison and number identification). The critical issue here is the amount variance accounted for in arithmetic skills at Time 2 after the effects of the control variables (earlier arithmetic skills, age, nonverbal abilities and vocabulary skills) had been accounted for. Table 3 shows the results of these analyses. Only letter comparison and number identification explained additional variance in arithmetic skills at Time 2 after the control variables had been entered into the equation (letter comparison 3.2%, and number identification 4.5%). Critically, magnitude comparison was not a unique predictor of arithmetic at Time 2 (and in fact it was a poorer predictor of later arithmetic skills than the letter judgement task which involved no access to numerical information). Only number identification is a unique predictor of arithmetic at Time 2 (accounting for 4% unique variance after all other predictors in the model were entered).

In a further model (Figure 2) we assessed whether earlier variations in arithmetic skills and other theoretically relevant variables are able to explain variations in magnitude comparison ability at Time 2. The data fitted the model well, *χ2* (163) = 220.59, *p* = .002, Root Mean Square Error of Approximation (RSMEA) = .045 (90% CI = .028-.060), Comparative Fit Index (CFI) = .962, Standardized Root Mean Residuals (SRMR) = .045. It is clear that only earlier magnitude comparison skills (the autoregressor) predicted later magnitude comparison skills. Fifty-four percent of the variance in magnitude comparison at Time 2 was explained in this model. These results are confirmed and extended by the results of the Cholesky models shown in Table 4: none of the predictor constructs explained variance in later magnitude comparison skills beyond earlier magnitude comparison skills. Thus there is no evidence of an influence of earlier arithmetic skills on the development of later magnitude comparison skills. The correlations between all latent variables in this model are shown in Table 2.

**Discussion**

This large-scale longitudinal study has explored some possible causal influences on the development of early arithmetic skills. Our particular focus was on the extent to which individual differences in the efficiency of an approximate number sense (ANS) system might place constraints on the development of arithmetic skills (Libertus et al., 2013; Libertus, Feigenson & Halberda, 2011; Piazza & Dehaene, 2004). To assess the ANS construct we used tasks that assessed the efficiency with which children could judge the relative numerosity of arrays of squares and the relative magnitudes represented by Arabic digits. The model shown in Figure 1 demonstrates clearly that these symbolic (numeral) and non-symbolic (arrays of squares) magnitude judgement tasks load onto a single factor (see also Kolkman, Kroesbergen & Leseman, 2013). The speed with which children performed these tasks is incompatible with the use of counting strategies to solve them and this, coupled with the correlations between these tasks, provides support for the engagement of a common numerical comparison process (which is partially distinct from the processes involved in the directly comparable letter comparison tasks).

Our second major finding is that the efficiency of the ANS at 6 years (indexed by our magnitude comparison latent variable) is not a unique predictor of arithmetic some 11 months later (although our ANS measure shows a strong longitudinal correlation with later arithmetic). This calls into question the widely held belief (e.g. Piazza, 2010) that the ANS places constraints on the development of arithmetic skills (at least in the period of development studied here). An important question for future longitudinal studies is whether variations in the efficiency of the ANS as measured here would be an important longitudinal predictor of later arithmetic skills if assessed in younger children (say 4 year olds).

To our surprise, number identification assessed at 6 years of age was a powerful independent predictor of arithmetic growth over the next 11 months. In this task children heard a spoken number and had to select the appropriate Arabic number from a choice of 4 or 5 options. This task measures at least two skills: Arabic digit knowledge and place-value understanding. We speculate that Arabic digit knowledge at school entry may be a key foundation for the development of later arithmetic skills (e.g. Kolkman et al., 2013; Krajewski & Schneider, 2009; Mundy & Gimore, 2009). Such an effect might be seen as directly analogous to the role of early letter knowledge as a critical longitudinal predictor of reading development (e.g. Caravolas et al., 2012; Hulme, Bowyer-Crane, Carroll, Duff, & Snowling, 2012). In short, for both arithmetic and reading development, we suggest that learning the symbol set (Arabic numerals or letters respectively) and their verbal labels is a critical foundational skill.

In addition to knowledge of single Arabic digits, an understanding of multi-digit numbers, and especially place-value understanding, is also crucial for arithmetic development. Moeller, Pixner, Zuber, Kaufmann and Nuerk (2011), for example, showed that place-value understanding in 7-year-old children predicted addition performance two years later. An understanding of place-value was also one component required for success on the number identification task used in this study.

The precision of the ANS system appears to increase with age (Halberda & Feigenson, 2008) and it has been suggested that such age related improvements may arise partly as a result of experience with formal arithmetic instruction (Piazza et al., 2013). To assess whether arithmetic skills play a role in refining the precision of the ANS we assessed predictors of change in magnitude comparison skills between Times 1 and 2 (see Figure 2). There was no evidence that arithmetic skill at time 1 predicted growth in the efficiency of magnitude comparison skills. Further studies over more extensive periods of development are needed to examine this issue further.

In summary, our results clarify the much debated role of the ANS in early arithmetic development. While it appears that an ANS system underpins the ability of children to judge the magnitudes represented by both arrays of squares and digits, contrary to some other claims (e.g. Libertus et al., 2011, Piazza, 2010; Piazza & Dehaene, 2004), we have found no evidence that this system places specific constraints on the development of arithmetic skills between the ages of 6 and 7 years. In contrast, the ability to match multi-digit Arabic numerals to their verbal labels is a strong predictor of individual differences in the growth of arithmetic in this developmental period. This latter finding suggests that verbal processes concerned with learning the labels for Arabic numerals, and the ability to translate between Arabic numerals and verbal codes, are critical for arithmetic development.

**References**

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition, 86(3)*, 201-221. doi:10.1016/S0010-0277(02)00178-6

Bugden, S., & Ansari, D. (2011). Individual differences in children’s mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition, 118(1),* 32-44. doi:10.1016/j.cognition.2010.09.005

Caravolas, M., Lervåg, A., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., ... & Hulme, C. (2012). Common patterns of prediction of literacy development in different alphabetic orthographies. *Psychological Science, 23(6)*, 678-686. doi: 10.1177/0956797611434536

[Dehaene](http://www.unicog.org/biblio/Author/DEHAENE-S.html), S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1- 42. doi: 10.1016/0010-0277(92)90049-N

De Smedt, B., Nöel, M-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behaviour. *Trends in Neuroscience and Education, 2,* 48-55. http://dx.doi.org/10.1016/j.tine.2013.06.001

Desoete, A., Ceulemans, A., De Weerdt, F., & Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non‐symbolic comparison tasks in kindergarten? Findings from a longitudinal study. *British Journal of Educational Psychology, 82(1),* 64-81. doi 10.1348/2044-8279.002002

Dunn, L. M., Dunn, D. M., & Styles, B. (2009). *British Picture Vocabulary Scale*, 3rd ed. London: GL Assessment.

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low‐income homes: contributions of inhibitory control. *Developmental Science, 16(1)*, 136-148. doi: 10.1111/desc.12013

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*(5), 1457-1465. doi: 0.1037/a0012682

Halberda, J., Mazzocco, M.M.M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*, 665‐668, doi:10.1038/nature07246

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

Hulme, C., Bowyer-Crane, C., Carroll, J. M., Duff, F. J., & Snowling, M. J. (2012). The Causal Role of Phoneme Awareness and Letter-Sound Knowledge in Learning to Read Combining Intervention Studies With Mediation Analyses. *Psychological Science, 23(6*), 572-577. doi: 10.1177/0956797611435921

Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction, 25*, 95-103. doi: 10.1016/j.learninstruc.2012.12.001

Krajewski, K., & Schneider, W. (2009). Exploring the Impact of Phonological Awareness, Visual-Spatial Working Memory, and Preschool Quantity--Number Competencies on Mathematics Achievement in Elementary School: Findings from a 3-year Longitudinal Study. *Journal of Experimental Child Psychology, 103(4),* 516-531. doi: 10.1016/j.jecp.2009.03.009

Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14(6)*, 1292-1300.doi: 10.1111/j.1467-7687.2011.01080.x

Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences, 25*, 126-133. doi: 10.1016/j.lindif.2013.02.001

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. doi:10.1371/journal.pone.0023749

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, H.-C. (2011). Early place-value understanding as a precursor for later arithmetic performance – a longitudinal study on numerical development. *Research in Developmental Disabilities, 32*, 1837-1851. doi: 10.1016/j.ridd.2011.03.012

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

Muthén, L. K., & Muthén, B. O. (1998–2013)*. Mplus user's guide. 7th edition.* Los Angeles, CA: Muthén & Muthén.

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences, 14(12)*, 542-552. doi: 10.1016/j.tics.2010.09.008

Piazza, M., & Dehaene, S. (2004). From number neurons to mental arithmetic: The cognitive neuroscience of number sense. In M. S. Gazzaniga (Ed.), *The Cognitive Neurosciences*, (3rd ed.) (pp. 865-77). Cambridge, MA: MIT Press.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., ... & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116(1)*, 33-41, doi: 10.1016/j.cognition.2010.03.012

Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychological Science, 24(6)*, 1037-1043. doi: 10.1177/0956797612464057

Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge: Mediation by numeral knowledge. *Journal of Educational Psychology, 105(2)*, 453-464. doi: 10.1037/a0031753

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. Oxford, England: Oxford Psychologists Press/San Antonio, TX: The Psychological Corporation.

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2012). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology. 114(3)*, 418-431. doi: 10.1016/j.jecp.2012.10.012

Wechsler, D. (2005). Wechsler Individual Achievement Test – Second UK Edition (WIAT-IIUK). London: Harcourt Assessment.

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

Table 1

*Means, Standard Deviations and Reliabilities for all Variables at all Time Points*

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  | Time 1 | |  | Time 2 | |
|  |  |  |  | M (sd) | Reliability |  | M (sd) | Reliability |
| Arithmetic | | |  | 3.57 (1.95) | .69 |  | 6.26 (2.55) | .75 |
| Magnitude comparison | | |  |  |  |  |  |  |
| Digit far | | |  | 20.54 (6.02) |  |  | 25.14 (5.16) |  |
| Digit close | | |  | 17.28 (5.68) |  |  | 20.97 (4.61) |  |
| NSF\_SS | | |  | 21.91 (6.68) |  |  | 28.27 (6.55) |  |
| NSC\_SS | | |  | 12.56 (4.71) |  |  | 17.16 (4.98) |  |
| 3\_4 SA | | |  | 13.17 (5.02) |  |  | 17.39 (4.85) |  |
| 5\_6 SA | | |  | 11.14 (4.38) |  |  | 14.45 (3.85) |  |
| 7\_8 SA | | |  | 8.73 (3.67) |  |  | 11.24 (3.62) |  |
| Letter comparison | | |  |  |  |  |  |  |
| Letter far | | |  | 8.59 (4.69) |  |  | – | – |
| Letter close | | |  | 5.16 (3.38) |  |  | – | – |
| Number ID | | |  | 5.08 (1.59) | .73 |  | – | – |
| BPVS | | |  | 30.86 (3.27) | .65 |  | – | – |
| Raven | | |  | 12.23 (4.11) | .73 |  | – | – |

Table 2

*Correlations between the Latent Variables in Figure 1 and Figure 2 Shown in the Upper and Lower Triangular Matrixes Respectively*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 Arithmetic T1 | | | – | .85\*\* | .48\*\* | .47\*\* | .43\*\* | .71\*\* | .17\* | .56\*\* |
| 2 Arithmetic T2 | | | – | – | .46\*\* | .59\*\* | .60\*\* | .77\*\* | .26\*\* | .60\*\* |
| 3 Vocabulary T1 | | | .48\*\* | – | – | .25\* | .28\* | .62\*\* | .31\*\* | .31\*\* |
| 4 Nonverbal abilities T1 | | | .46\*\* | – | .25\* | – | .54\*\* | .42\*\* | .18\*\* | .45\*\* |
| 5 Letter comparison T1 | | | .41\*\* | – | .26\* | .54\*\* | – | .40\*\* | .23\*\* | .66\*\* |
| 6 Number ID T1 | | | .70\*\* | – | .62\*\* | .41\*\* | .38\*\* | – | .20\* | .45\*\* |
| 7 Age | | | .16 | – | .31\*\* | .18\* | .23\*\* | .20\* | – | .29\*\* |
| 8 Magnitude comparison T1 | | | .56\*\* | – | .31\*\* | .45\*\* | .65\*\* | .44\*\* | .28\* | – |
| 9 Magnitude comparison T2 | | | .52\*\* | – | .37\*\* | .41\*\* | .46\*\* | .49\*\* | .20\*\* | 70\*\* |

*Note*. \* = *p* < .05. \*\* = *p* < .01.

Table 3

*Cholesky Model Predicting Arithmetic at Time 2*

|  |  |  |
| --- | --- | --- |
| Step | T1 Predictors: | Arithmetic T2 |
|  |  |  |
| 1 | Age | .067\*\* |
| 2 | Arithmetic T1 | .671\*\* |
| 3 | Raven | .043\* |
| 4 | BPVS | .001 |
|  |  |  |
| 5 | Number ID | .045\* |
| 6 | Magnitude comparison | .005 |
| 6 | Letter comparison | .027\* |
|  |  |  |
| 5 | Magnitude comparison | .006 |
| 6 | Number ID | .044\* |
| 6 | Letter comparison | .027 |
|  |  |  |
| 5 | Letter comparison | .032\* |
| 6 | Number ID | .040\* |
| 6 | Magnitude comparison | .000 |
|  |  |  |
| 7 | Number ID | .040\* |
| 7 | Magnitude comparison | .000 |
| 7 | Letter comparison | .023 |
|  |  |  |
| *R2* |  | .855\*\* |

Note. All values are squared beta coefficients; \* = *p* < .05. \*\* = *p* < .01.

Table 4

*Cholesky Model Predicting Number Comparison at Time 2*

|  |  |  |  |
| --- | --- | --- | --- |
| Step | T1 Predictors: | Magnitude Comparison T2 | |
|  |  |  | |
| 1 | Age | .041\* | |
| 2 | Magnitude Comparison T1 | .452\*\* | |
| 3 | Raven | .010 | |
| 4 | BPVS | .025\* | |
|  |  |  | |
| 5 | Number ID | .014 |
| 6 | Arithmetic | .000 |
| 6 | Letter comparison | .0021 |
|  |  |  |
| 5 | Arithmetic | .006 |
| 6 | Number ID | .009 |
| 6 | Letter comparison | .0011 |
|  |  |  |
| 5 | Letter comparison | .0021 |
| 6 | Number ID | .014 |
| 6 | Arithmetic | .005 |
|  |  |  | |
| 7 | Number ID | .009 | |
| 7 | Arithmetic | .000 | |
| 7 | Letter comparison | .0021 | |
|  |  |  | |
| *R2* |  | .543\*\* | |

Note. All values are squared beta coefficients; 1 = Negative beta value; \*\* = *p* < .01.

**Figure Legends**

*Figure 1*. Arithmetic at Time 2 predicted by all constructs measured at Time 1. Ellipses reflect latent variables and rectangles reflect observed variables. One-headed arrows from the latent to the observed variables reflect factor loadings in the measurement model and one-headed arrows between the latent variables reflect true-score regressions between constructs. The one-headed arrow from a number into the latent variable reflects the residual of the construct. Two-headed arrows between the latent variables reflect true-score correlations between constructs. Solid lines indicate statistically significant relationships and dotted lines indicate statistically non-significant relationships.

*Figure 2.* Magnitude comparison at Time 2 predicted by all constructs measured at Time 1. Ellipses reflect latent variables and rectangles reflect observed variables. One-headed arrows from the latent to the observed variables reflect factor loadings in the measurement model and one-headed arrows between the latent variables reflect true-score regressions between constructs. The one-headed arrow from a number into the latent variable reflects the residual of the construct. Two-headed arrows between the latent variables reflect true-score correlations between constructs. Solid lines indicate statistically significant relationships and dotted lines indicate statistically non-significant relationships.