Article:
Habermann, Stefanie, Donlan, Chris, Göbel, Silke M orcid.org/0000-0001-8845-6026 et al. (1 more author) (2020) The critical role of Arabic numeral knowledge as a longitudinal predictor of arithmetic development. Journal of Experimental Child Psychology. p. 104794. ISSN: 0022-0965

https://doi.org/10.1016/j.jecp.2019.104794

The critical role of Arabic numeral knowledge as a longitudinal predictor of arithmetic development.

.

Stefanie Habermann*, Chris Donlan*, Silke M. Göbel** & Charles Hulme***

* Department of Language and Cognition, University College London, London, UK

**Department of Psychology, University of York, UK

***Department of Education, University of Oxford, UK

Acknowledgements: We are grateful to all the children and school staff who contributed so generously to this research. The study was supported by an ESRC studentship awarded to the first author.

The critical role of Arabic numeral knowledge as a longitudinal predictor of arithmetic skills.

Abstract

Understanding the cognitive underpinnings of children's arithmetic development has great theoretical and educational importance. Recent research suggests symbolic and nonsymbolic representations of number influence arithmetic development before and after school entry. We assessed non-verbal ability and general language skills, as well as nonsymbolic (numerosity) and symbolic (numeral) comparison skills, counting and Arabic numeral knowledge (numeral reading, writing and identification) in preschool children (age 4 years). At age 6 we re-assessed nonsymbolic (numerosity) and symbolic (numeral) comparison and arithmetic. A latent variable path model showed that Arabic numeral knowledge (defined by numeral reading, writing and identification at age 4 years) was the sole unique predictor of arithmetic at age 6 years. We conclude that knowledge of the association between spoken and Arabic numerals is one critical foundation for the development of formal arithmetic.

Keywords: Arabic numeral knowledge; Magnitude comparison; Pre-school; Formal arithmetic; Approximate number system

Introduction

Understanding the cognitive skills that form the foundation of children's arithmetic development is critical for theory and for educational practice. While domain-general factors such as language abilities (LeFevre et al., 2010), working memory (De Smedt, Janssen et al., 2009: Simanowski & Krajewski, 2017) and executive functions (Bull & Scerif, 2001; Cragg, Keeble, Richardson, Roome, & Gilmore, 2017) are established predictors of arithmetic development, there is also substantial evidence for the importance of domain-specific factors, including magnitude representation (Mazzocco, Feigenson, & Halberda, 2011; Xenidou-Dervou, Molenaar, Ansari, van der Schoot, & van Lieshout, 2017), counting (Cirino, 2011; Chu, vanMarle, Rouder & Geary, 2018) and number knowledge (Göbel, Watson, Lervåg & Hulme, 2014). The developmental influence of these domain-specific factors is not yet fully understood.

The proposal that there is a preverbal approximate number system (ANS) that we share with non-human animals (Nieder & Dehaene, 2009; Piazza, 2010), is supported by evidence from studies of infants. Particular interest has been attached to the developmental importance of nonsymbolic numerosity comparison tasks as a measure of the precision of the ANS. There is evidence (Mazzocco, Feigenson, & Halberda, 2011) that performance on nonsymbolic numerosity comparison in preschool predicts later arithmetic. Several meta-analyses (e.g.; Fazio, Bailey, Thompson, & Siegler, 2014; Schneider et al., 2017) report significant longitudinal and concurrent relationships between performance on nonsymbolic numerosity comparison tasks and arithmetic. This relationship, however, is significantly weaker than the relationship between symbolic numeral comparison and arithmetic performance reported in the same meta-analyses (De Smedt, Noël, Gilmore, & Ansari, 2013; Fazio et al., 2014; Schneider et al., 2017). In a recent study comparing the predictive power of nonsymbolic numerosity comparison and symbolic numeral comparison directly, symbolic numeral comparison in primary school was the stronger predictor of children's future arithmetic performance (Xenidou-Dervou et al., 2017).

A new perspective on the relation between nonsymbolic numerosity comparison and symbolic numeral comparison and arithmetic was provided by Göbel et al. (2014). Six year olds' cognitive and numerical skills, including multiple measures of nonsymbolic numerosity comparison and symbolic numeral comparison, were assessed. These were shown to comprise a unitary 'magnitude comparison' factor. The study also included a number identification task in which spoken numerals had to be matched to the corresponding Arabic numeral. Despite a strong longitudinal correlation between magnitude comparison and arithmetic skills measured 11 months later (r=.60), a path model showed the contribution of magnitude comparison to be completely subsumed by number identification which, apart from the autoregressive effect of earlier arithmetic, was the only significant predictor of arithmetic skill at age 7 years, accounting for 32% of variance in outcome.

Göbel et al. (2014) interpret their number-identification task as tapping individual differences in both Arabic numeral knowledge and place-value understanding, suggesting that the former may represent a critical foundational skill underlying early arithmetic, analogous to the role of letter knowledge in reading. These findings clarify the previous literature as follows. On the one hand, consistent findings of high correlations between single-digit symbolic numeral comparison and nonsymbolic numerosity comparison (Holloway & Ansari 2009; Göbel et al., 2014; Matejko & Ansari 2016) may reflect general properties of the mental representation of magnitude, relevant but not central to arithmetic development. On the other hand, the specific relation between single-digit symbolic numeral comparison and early arithmetic skills (Holloway & Ansari, 2009; Vanbinst, Ansari, Ghesquière & De Smedt, 2016) may reflect the contribution of symbolic item identification as a foundational arithmetic skill. However, single-digit symbolic numeral comparison is limited in range, and may fail to capture the precision and extent of symbol knowledge (especially the understanding of place-value) needed to drive arithmetic development. Thus, when a comprehensive (multi-digit) numeral identification task is included in a longitudinal model of early arithmetic development, as in the study by Göbel et al. (2014), magnitude comparison (symbolic or nonsymbolic) does not account for any additional variance in arithmetic. According to this account, the correspondence between symbolic and nonsymbolic comparison is based on a shared quantity representation. In contrast, the correspondence between single-digit symbolic numeral comparison and arithmetic is based on (limited) knowledge of the symbol system, and is subsumed by the correspondence between multi-digit item identification and arithmetic, which requires more advanced knowledge of the symbol system.

Central to our argument here is the claim that learning the correspondence between spoken and written multi-digit numerals provides a pathway to understanding the number system, which is fundamental to performing mathematical operations such as addition and subtraction. In some ways this is a counterintuitive claim. Much previous research invokes the notion that nonsymbolic representations of magnitude underpin the development of number knowledge, whether via a nonsymbolic ANS (Halberda & Feigenson 2008; Mazzocco et al. 2011), or through the association between single-digit numerals and the magnitudes they represent (Holloway & Ansari 2009; Schneider et al. 2017). An alternative account, referring particularly to multi-digit numbers, has been elaborated in a recent study by Yuan, Prather, Mix & Smith (2019). This builds on earlier work (Mix et al. 2014) which showed that 3- and 4-year-old children acquire significant knowledge of the correspondence between spoken and written multi-digit numerals. Yuan et al., (2019) replicate this finding in a large and diverse sample across a wider age range, and go on to demonstrate that multi-digit spoken/written numeral mapping substantially precedes the ability to map multi-digit spoken numerals to corresponding quantities. Furthermore, there is no prediction, beyond what is accounted for by age, from quantitative mapping to symbolic mapping, or vice versa. Yuan et al. (2019) interpret their findings by reference to the triple code model (Dehaene, 1992), suggesting that the development of the triple code system is likely to be more complex than previously thought. While knowledge of the association between single-digit numbers and the quantities they represent is likely to provide an essential foundation for learning about multi-digit numbers, it does not follow that this framework of association is necessarily extended to include the much more extensive correspondence between quantities and multi-digit numbers. The recursive property of multi-digit numerals is captured more effectively by an alternative mechanism which capitalizes on their linguistic representation. Decade-unit relations are explicit in the spoken forms e.g. *twenty-one* to *twenty-nine*, *thirty-one* to *thirty-nine*, etcetera. Further clarity is evident in the hundreds, where the single digit forms are fully preserved, *one hundred*, *two hundred*, etcetera. While the intrinsic structure of multi-digit numbers is more transparently represented in other languages (Miura & Okamoto, 1989), English spoken numbers show broad consistency with the structure of Arabic numerals. Thus it is plausible to argue that, even for English speakers, learning about multi-digit numbers does not begin with mapping of specific numbers to the quantities they represent, but rather with learning about the relational patterns which hold between spoken and written forms.

An important step in understanding the implications of early multi-digit number knowledge was provided by Goebel et al. (2014), who showed for the first time that the ability to associate spoken and written multi-digit numerals predicts early arithmetic knowledge, over and above single-digit magnitude comparison. However, these data were collected while children were in school where teaching about multi-digit numbers, as well as simple arithmetic, is mandatory. Thus it is possible that these particular inputs are reflected in the study findings. A stronger test of the multi-digit hypothesis (implied but not evaluated by

Yuan et al. 2019) requires a longitudinal study spanning pre-school (where multi-digit knowledge is not taught) to the early years of formal schooling, employing a comprehensive range of predictors including nonsymbolic and single-digit magnitude comparison.

A further strand of research has examined the importance of children's counting skills. Using the Give-a-Number task devised by Wynn (1990), a recent study by Chu, vanMarle et al. (2018) identified object counting skills in 3 and 4-year-olds as the primary predictor of the strategies (e.g. counting, decomposition, retrieval) they use to solve arithmetic problems three years later. It is not clear whether rote-counting (simple production of the count word sequence) had been included as a potential pre-school predictor in this analysis of strategies, but rote-counting was found to be a mediator of long-term outcome (arithmetic strategy at age 7). Naming Arabic digits (in the range 1-15) was included in both the initial test battery and the predictive model, and was found to be an independent predictor of arithmetic, alongside object counting. Cirino (2011), studying 6-year-olds in kindergarten, found that both rote-counting and object counting were significant independent concurrent predictors of addition skill, alongside number knowledge and symbolic comparison. Koponen, Salmi & Eklund (2013) also studied 6-year-olds in pre-school, and found that rote-counting skill at that age significantly predicted arithmetic fluency at age 10, while a pre-school number concept task similar to the Give-a-number task was not predictive.

The findings reported above, and others (e.g., Lyons et al., 2014), highlight the potentially time-sensitive nature of influences on arithmetic development. Might it be that the findings of Göbel at al. (2014) represent a transient state in which symbol identification has particular importance? Would symbolic numeral knowledge (beyond single-digit comparison) assessed in pre-school be a powerful predictor of later arithmetic? Or would early measures of the precision of the ANS, measured before the onset of formal schooling, be a better predictor?

Unique amongst the predictor variables identified in the studies cited above, the identification of multi-digit symbols, as implemented by Göbel et al. (2014), explicitly requires transcoding from spoken to written numeral forms. This is a language-specific task which, in English and many other Western languages, entails mastery of systematic but inconsistent correspondences between Arabic numerals and corresponding number words (e.g. 14 maps on to 'fourteen' not on to 'ten four'). An influential study of 7-year-olds writing multi-digit numbers to dictation (Dal Martello, 1990) identified syntactic errors (errors of place value, e.g., writing 1006 for 'one hundred and six') as predominant. Thus the simple (but not trivial) challenge of learning arbitrary associations between spoken and Arabic single digit numerals is complicated significantly for multi-digit numbers which require a proper understanding of place value. The processes by which this learning is achieved remain largely unexplored. We know that the ability of English-learners to recite the spoken numeral sequence up to 'ten' is typically established by age 4 years, and that mapping of these spoken numerals to the quantities they represent may take place by age 5 years (Le Corre & Carey, 2007). High accuracy in symbolic number comparison, indicating mapping from Arabic numerals to their corresponding magnitudes, is recorded in 5 year olds for numerals 1-5, in 6 year olds for numerals 1-9 and in 7 year olds for double digit numerals (Donlan, Bishop & Hitch, 1998; Sekuler and Mierkiewicz, 1977; Lyons, Price et al., 2014). However, a growing body of evidence (Mix, Prather, Smith and Stockton, 2014; Yuan et al. 2018) indicates that 3- and 4-year-olds have substantial knowledge of the relation between the spoken and written forms of multi-digit numerals. In addition, learning this complex interrelation precede, and are largely independent of, the mapping from spoken to quantitative (nonsymbolic) representations (Yuan et al. 2018). Evidence from Chu et al. (2018), cited above, indicates that 3 to 4-year-olds' transcoding skills (naming Arabic digits), significantly influence arithmetic strategies three years later, independent of the effect of object counting. Also notable in Chu et al. (2018) is the role of rote-counting (simple recitation of the spoken sequence) as an influential factor in early arithmetic development in its own right (see also Koponen, Aunola, Ahonen & Nurmi, 2007; Donlan, Cowan, Newton & Lloyd 2007; Koponen et al. 2013). Thus a range of studies, varying in age and educational settings, indicate the importance of different aspects of symbolic number knowledge, over and above the mapping from single digit numerals to magnitudes, as precursors to arithmetic skills.

Direct instruction in arithmetic typically begins at school entry . International variation in the age at which children start school may therefore influence the factors affecting mathematical development. Researchers seeking to map the precursors of arithmetic skill may select samples differing widely in age, according to the education systems within which they work. In Finland, where formal schooling starts when children are seven, Koponen et al. (2013) assessed basic skills at kindergarten entry (mean age 6:02), and tested outcomes (reading and arithmetic) at ages nine and ten. In Canada, where formal schooling starts when children are five or six, LeFevre et al. (2010) recruited a combined group of pre-schoolers (at median age 5;0) and kindergartners (at mean age 5;11), and followed them up two years later. Studies carried out in the US, where formal school starts when children are six, show a similar sampling pattern. Cirino (2010), in a cross-sectional study of precursors of arithmetic, assessed kindergartners at mean age 6.02. Likewise, in the Netherlands, Xenidou-Dervidou et al. (2018) assessed participants early in their kindergarten year (mean age 5;09), with subsequent follow-up in Grades 1, 2 and 3. In contrast to all the previous examples, the education system in England requires formal schooling for children when they reach 5 years of age, a year earlier than in many other countries. Therefore the sample first tested by Göbel at al. (2014) at mean age 6;03, had entered school up to a year earlier, and received direct instruction in arithmetic during that time. Given these variations, it is possible that the predictors of arithmetic development across different studies conducted in different countries reflect at least in part differences in both the teaching received in school and the chronological age at which children were assessed. If formal instruction is provided at 5, 6 or 7 years of age, then the notion of a precursor skill may need to be redefined accordingly. Further complexity is added at the pre-school level. The study by Chu et al. (2018) provides an example. Participants were recruited from a federally funded pre-school programme targeting low-income families in the US. Thus, instead of entering the school system at age 5 (kindergarten), these children entered pre-school and were first assessed at mean age 3;10, were assessed again in their second pre-school year at mean age 5;0, entered their kindergarten year and were assessed once more at mean age 5;10, with final follow-up in first grade (formal school) at mean age 6;9.

The present study focusses particularly on the nature and extent of children's Arabic numeral knowledge, preceding their exposure to formal arithmetic, and on the strength of prediction from different aspects of symbolic number knowledge at age 4 years to arithmetic skill two years later, after one year of formal schooling. Unlike previous studies of this age group we emphasise the relation between spoken number words and Arabic numerals. As well as assessing writing to dictation (spoken input, written output) we assess numeral naming (reading printed Arabic forms) and numeral identification (matching spoken input to printed Arabic numerals). Following Mix et al. (2014), the range of numerals tested includes two- and three-digit numbers. We examine these multiple measures of 'Arabic numeral knowledge', alongside rote-counting (simple production of the spoken number sequence), single digit symbolic numeral comparison and nonsymbolic numerosity comparison, and also include general measures of language and non-verbal ability in order to differentiate the number-specific content of our predictors. We use confirmatory factor analysis to construct a model of the structure of number skills and number knowledge at age 4 years. We assess the measures taken at age 4 years as predictors of arithmetic skill measured two years later at age 6 years.

Our study asks the following questions:

1. Do pre-school children's nonsymbolic numerosity comparison and symbolic numeral comparison form a unitary factor, and is there change over time in their association?
2. To what extent are pre-school (age 4 years) measures of nonsymbolic numerosity comparison, symbolic numeral comparison, Arabic numeral knowledge, and counting unique predictors for arithmetic skills at age 6 years?

Methods

Participants

Typically developing children in one large combined nursery and primary school were assessed. To establish socioeconomic status, the English Indices of Deprivation (UK Government Statistics, 2015) were used. The overall Index of Multiple Deprivation of the school catchment area indicates that this area shows less than average deprivation (rank 25,631 out of 32,844, where 1 is most deprived), corresponding to a middle to high socio-economic background.

One hundred children (48 boys and 52 girls, mean age = 4 years 2 months (*SD* = 3.5 months)) participated at Time 1 (summer term of pre-school year), 75 children (mean age = 4 years 11 months (*SD* = 3.5 months)) were reassessed at Time 2 (autumn term of formal school reception year, approximately 9 months later) and 71 children (mean age = 6 years 4 months (*SD* = 3.5 months) were reassessed at Time 3 (summer term of Year 1 of formal schooling, approximately 25 months later).

Time 1 Measures

The tests reported here were part of a larger test battery. All tests were divided into counterbalanced sessions of 20 to 40 minutes. All tests were individually administered in a separate room or another quiet place in the school.

Nonverbal intelligence. Nonverbal intelligence was assessed using set A of the Raven's Coloured Progressive Matrices (Raven's CPM; Raven, Court, and Raven, 1993) was chosen. Items were administered according to the manual. Children were given an incomplete matrix puzzle and asked to choose the piece that completes the matrix. Before testing, three novel practice items were administered. These were created based on the features of the original matrices. One point was given for each correct response to a test item, with a maximum possible score of 12.

Grammatical ability. The children's grammatical ability was assessed using the Test for Reception of Grammar II (TROG-2; Bishop, 2003). The raw scores (number of blocks passed) were recorded (maximum score = 20).

Vocabulary knowledge. Children's vocabulary skills were examined using the British Picture Vocabulary Scale 3rd Edition (BPVS - III; Dunn, Dunn and Styles, 2010). The raw scores (number of correct responses) were reported (maximum score = 168).

Numeral Identification. Children were presented with four Arabic numerals and were asked to point to the numeral that matched the verbally presented target number. Distractors were chosen on the basis of common errors with place-value and visual similarity to the targets (e.g., for the target number "206," choices were 206, 260, 26, 2060). Target numbers were 6, 28, 206, 7, 91, 2, 41, 52, 11, 69, 37, 43, 74, 168, 13 and 85 (maximum score = 16). Full details of the stimuli for this task are shown in the Appendix.

Numeral Writing. To assess children's number-writing skills, we asked them to transcribe twelve Arabic numerals (2, 9, 7, 4, 8, 10, 6, 1, 20, 3, 100 and 5) which were presented in spoken form. Two points were awarded for each numeral (accuracy and orientation, scoring was based on the Letter Writing task of Caravolas, et al. 2012; maximum score = 24).

Numeral Reading. Children were asked to read aloud the Arabic numerals (MS Office 2013, Comic Sans MS, size 350) one to ten, presented in random order. One point was awarded for each correct number (maximum score = 10).

Rote Counting. Children were asked to count aloud from one. Testing stopped after the child reached the number 111 or did not know how to count further. The highest number counted without mistakes was reported (maximum score = 111).

Symbolic Numeral Comparison. Pairs of Arabic numerals were displayed within two adjacent boxes (12cm x 12cm) with digits in Calibri font size 350. Digits ranged from one to nine and both orders of the pairs were presented (e.g. 3 and 4, 4 and 3). To investigate the numerical distance effect (Moyer and Landauer, 1967) two versions were administered. In the close version the difference between the two digits was one or two and in the far version the difference was five, six or seven. Each digit pair was presented on a single page. Children were given one point for every correct response with a maximum score of 32 (16 close items and 16 far items).

Nonsymbolic Numerosity Comparison. Tasks were based on those of Göbel et al. (2014). Arrays of black squares were presented within 12cm x 12cm boxes.. The size of the squares in the arrays was manipulated in two conditions: *fixed size* (squares in each box had a fixed size) and *surface-area matched* conditions (the size of the squares was matched for total surface area in black; smaller numerosities had larger squares and larger numerosities had smaller squares, but overall area in black was the same).In the *fixed size* condition, numerosities presented ranged between 5 and 13. For *close* trials in the fixed condition the difference between arrays was one or two squares, and for *far* trials the difference between arrays was five, six or seven. The *surface-area matched* condition examined larger numerosities ranging from 20 to 40. Given these numerosities, comparisons were based on ratio rather than difference between stimuli. Similar to Göbel et al. (2014), baseline numbers 20 through to 30 were compared to their nearest whole number of the ratios 2:3 and 3:4, e.g. 23 was compared to 35 (2:3) and 31 (3:4). The numerosity comparison tasks were individually administered, blocked according to subtasks with one block containing the *surface-area matched* items and the other containing the *fixed size* items. The subtasks were presented along with the other numerical tasks, with the order of tasks randomized across participants. Each subtask has 16 trials.

Time 2 Measures

Arithmetic. Children's basic arithmetic skills were assessed at Time 2 using simple addition problems. The test comprised of two parts matched for difficulty with ten simple additions with sums less than ten in each part. Children were given three minutes per part to solve as many problems as possible. All arithmetic problems were presented in Arabic notation (Comic Sans MS, size 260pt) and, simultaneously, in the more familiar spoken format to the child. Problems were arranged so that additions with same sums or similar summands were never adjacent. Children were encouraged to use wooden sticks provided or their fingers if needed (maximum score = 10 per part).

Time 3

Nonsymbolic Numerosity Comparison and Symbolic Number Comparison. Assessments were conducted in a small group setting (5 to 8 children). The subtasks used at Time 1 were redesigned as a group test using the same stimulus pairs. All comparisons were presented in pairs of two adjacent 2.1 cm x 2.1 cm boxes. The subtasks were blocked within two booklets which were matched for difficulty level, and which included further subtasks not reported in the current study. *Fixed size* and *surface-area matched* subtasks were alternated within the booklets. The order of presentation of booklets was counterbalanced across participants. The order of target locations (left array vs. right array) was controlled in order to avoid repeated response patterns. Each subtask comprised of 36 item pairs. Six of the pairs were displayed on each page and there were six pages per subtask. Children were asked to tick the bigger number or box with more squares. Two practice trials were displayed on the first page of each subtask. The first trial was demonstrated by the experimenter who then asked children to tick the next box. Another six practice items were then given to the children, but only for the first three subtasks of the first booklet. Feedback was given on practice items but not on test trials. Children had 30 seconds per subtask to solve as many comparisons as possible.

Arithmetic. Assessment was carried out individually, using the Numerical Operations subtest of the second edition of the Wechsler Individual Achievement Test (WIAT-II; Wechsler, 2005). The first six items (identifying and writing Arabic numerals) were excluded for the present study in order to provide a more conventional measure of arithmetic, and to avoid the possible confound with number knowledge. The test was executed according to the manual and children were allowed to complete the task in their own time (maximum score = 25).

As a second measure of arithmetic, children's speeded arithmetic skills (fluency) was assessed using the following subtests of the Test of Basic Arithmetic and Numeracy Skills

(TOBANS; Brigstocke, Moll & Hulme, 2016): “addition” (single digit augend/addend, sum less than 10), “addition with carry” (single digit augend/addend, sum range 11-18), “subtraction” (single digit minuend/subtrahend, difference range 1-7). Children were asked to complete as many arithmetic problems as possible in one minute. One point was awarded per correct answer, even if the numeral was written backwards (maximum score addition/subtraction= 90; maximum score addition with carry = 30).

Results

In order to examine possible effects of attrition to the sample, we examined differences between participants who were assessed throughout the longitudinal study and those who only assessed at Time 1. Descriptive statistics and tests of differences between the groups are shown in Table 1. There were no significant differences between groups. We investigated this issue further by conducting Little's MCAR test on all measures at Time 1, Time 2 and Time 3. The test showed Chi-Square = 103.821, df = 99, p = .350, confirming that there is no evidence that missing data were not Missing Completely At Random (MCAR); i.e. there is no evidence that attrition was not random. All subsequent analyses were carried out on the whole sample with missing data being handled by Full Information Maximum Likelihood estimators. The means, standard deviations and reliabilities for all measures at all three time points are shown in Table 2. Correlations between all measures are shown in Table 3. Structural equation models were constructed using Mplus Version 7 (Muthén & Muthén, 2013).

Development of magnitude comparison: a two-factor model vs. a unitary model

A set of confirmatory factor analyses examined the structure of the magnitude comparison tasks at Time 1. Although all tasks loaded on a single magnitude factor (Figure 2), the one-factor model did not provide an acceptable fit to the data, RMSEA = .136 (90% CI = .072 - .203), CFI = .827, SRMR = .071. In contrast, as shown Figure 1, a two-factor model provided a good fit to the data, RMSEA = .058 (90% CI = .000 - .144), CFI = .973, SRMR = .047.

The second set of confirmatory factor analyses (Figure 2) analysed the structure of the magnitude comparison tasks at Time 3 (25 months later) and showed that a one-factor model provided a good fit to the data, RMSEA = .064 (90% CI = .000 - .132), CFI = .991, SRMR

= .023) so in contrast to Time 1, magnitude comparison at Time 3 appears to reflect a single unitary factor.

Longitudinal prediction of early arithmetic skills

A latent variable path model was used to investigate the longitudinal predictors of children's arithmetic skills at Time 2 and Time 3 (9 and 25 months later) from our Time 1 measures. The Time 1 latent variables were : nonverbal intelligence, general language comprehension, rote counting skills, Arabic numeral knowledge, symbolic and nonsymbolic comparison. The latent variables for nonverbal intelligence and counting were each defined by just one indicator by constraining the error variance to 1 minus the reliability of the test. In an initial version of this model, arithmetic at Time 2 was regressed on the Time 1 predictors and Time 3 arithmetic was regressed on the same time predictors as well as Time 2 arithmetic (see Figure 3). Nonsignificiant paths were trimmed iteratively, while checking that deleting nonsignificant paths resulted in no significant change in model fit. In the trimmed model shown in Figure 3 all retained paths are statistically significant

The model provides a good fit to the data, $\chi 2$ (134) = 168.331, p = .024, root-mean-square error of approximation (RMSEA) = .042 (90% CI = .016 - .060), comparative fit index (CFI) = .95, standardised root mean residual (SRMR) = .084. Number knowledge is a unique predictor of children's arithmetic skills at Time 2, accounting for 49 % of variance. Moreover, number knowledge was also the only unique longitudinal predictor of arithmetic scores at Time 3 (25 months after initial testing) accounting for 43 % of variance.

The pattern of correlations between the latent constructs in the model is shown in Table 3. Arabic numeral knowledge at Time 1 had the strongest correlation with arithmetic at both Time 2 and Time 3 (r=.67 and r=.64 respectively). Arabic numeral knowledge also correlated highly with language, counting and symbolic numeral comparison, whereas its relation to nonsymbolic numerosity comparison was relatively weak. Particularly notable is the strong association between language and both Arabic numeral knowledge (r=.51) and symbolic comparison (r=.62). These associations provide support for the independent Linguistic Pathway identified by LeFevre et al. (2010). In general, nonsymbolic numerosity comparison was less strongly associated with domain-specific measures than symbolic numeral comparison.

At first sight, it is surprising that in this model Time 2 arithmetic (the autoregressor) is not a unique predictor of Time 3 arithmetic. As shown in Table 3 the correlation between the latent variables for arithmetic at Time 2 and Time 3 is substantial (r= .47). The low path coefficient (Time 2 Arithmetic -> Time 3 Arithmetic) reflects the strong relationship between Time 2 Arithmetic and Time 1 Arabic Numeral knowledge (r=.67). Surprisingly the correlation between Time 1 Arabic Numeral knowledge and Time 3 Arithmetic (r=.64) is actually higher than the correlation between Time 2 Arithmetic and Time 3 Arithmetic. This pattern explains the very low path coefficient for Time 2 Arithmetic -> Time 3 Arithmetic (which is the partial regression coefficient controlling for the effects of Time 1 Arabic Numeral Knowledge).

Discussion

We set out to assess the role of Arabic numeral knowledge, counting, nonsymbolic numerosity comparison and symbolic numeral comparison at age 4 years as predictors of arithmetic skills at age 6 years (after one year of formal schooling). We also examined the factor structure of magnitude comparison at ages 4 and 6 years.

Our measures of Arabic numeral knowledge (writing numerals to dictation, numeral reading, numeral identification) defined a single latent factor. Confirming the findings of Mix et al. (2014), a substantial proportion of children were successful in processing multi-digit stimuli.

Our main focus is on pre-school predictors of arithmetic skills at age 6 years. The path model shown in Figure 3 provided an excellent fit to the data, and demonstrates that Arabic numeral knowledge at age 4 is the sole unique predictor of arithmetic skills at age 6, accounting for 64% of the variance. Arabic numeral knowledge was also a strong predictor of arithmetic at age 5.

Contrary to expectations based on previous studies (e.g., Cirino 2011; Koponen et al., 2013), there was no independent contribution of rote counting to later arithmetic skills. Our aim was to distinguish clearly between knowledge of the spoken number sequence on the one hand, and the relation between spoken and Arabic numerals on the other. Counting accounted for no unique variance in arithmetic after the effects of Arabic numeral knowledge were accounted for; this reflects the fact that these two variables share substantial common variance (r=.66) and that Arabic numeral knowledge was the more powerful predictor of later arithmetic. Likewise, although symbolic numeral comparison was more highly correlated with later arithmetic than nonsymbolic numerosity comparison (consistent with numerous previous findings and meta-analyses e.g. De Smedt et al., 2013; Schneider et al., 2017), neither of these factors predicted arithmetic after accounting for individual differences in Arabic numeral knowledge.

Our findings extend those of Göbel et al. (2014). We assessed children at an earlier developmental and educational stage, spanning the crucial transition from pre-school to early formal schooling. We also took more comprehensive measures of number knowledge, including writing Arabic numerals to dictation. Perhaps surprisingly, despite the earlier age studied, and the qualitative difference in educational experience, our findings mirrored those of Göbel et al. (2014) insofar as we found Arabic numeral knowledge to be the sole longitudinal predictor of later arithmetic skills.

Our finding of a change in the factor structure of magnitude comparison over the period of observation allows us to identify key elements in the development of symbolic representation of number preceding formal school entry. We found that symbolic numeral comparison becomes consolidated within a general magnitude comparison factor over the pre-school to formal school transition. The strong correlations we find between latent factors suggest shared involvement of rote counting, symbolic numeral comparison and Arabic numeral knowledge in the enhancement of symbolic representation of number over this time period. However, our results are unequivocal in suggesting that individual differences in the extent to which spoken and Arabic numeral forms are associated at age 4 (pre-school), is a very important factor in predicting arithmetic skills at age 6 (after one year of formal school).

LeFevre et al. (2010) proposed linguistic, quantitative and spatial attention factors measured at pre-school as independent precursors of later arithmetic skills. Their linguistic factor was based on a combination of number-specific and general language tasks. The results of the current study are compatible with those LeFevre et al. (2010) insofar as our predictive model attributes particular importance to the pathway from pre-school language-specific number knowledge to later arithmetic, and is based on strong associations between language and symbolic number processing factors.

The current findings, together with those of Göbel et al. (2014), invite re-examination of the widely-held view that symbolic number skills as assessed by single digit symbolic numeral comparison provide the basis for arithmetic development in the early years of schooling (Lyons et al., 2014; De Smedt et al., 2009, 2013; Vanbinst et al., 2016). In both the current study and Göbel et al. (2014) the longitudinal correlation between magnitude comparison and arithmetic is completely subsumed by measures of Arabic numeral knowledge. Thus we propose that the enhancement of number representation through the association of spoken and Arabic numerals has primacy over simple magnitude representation in the developmental process. In this account, early understanding of the complex structural correspondences between spoken and Arabic multi-digit numerals provides a critical foundation for learning formal (written) arithmetic skills.

The factor structure of magnitude comparison changed over the course of our longitudinal study. At age 4, a two-factor model distinguishing symbolic from nonsymbolic tasks showed a significant advantage over a single-factor model. At this time of measurement 31% of participants were at or below chance level in symbolic numeral comparison, and the majority of these individuals failed to name the digits 1-9 correctly. The final one-factor model (at age 6) is consistent with the findings of Göbel at al. (2014), indicating a common magnitude comparison process, serving both symbolic and nonsymbolic stimuli. A possible interpretation of the developmental process here is that both magnitudes and Arabic digits become strongly associated with verbal labels during this developmental period, meaning that comparisons between both digits and nonverbal magnitudes come to depend upon a common system in which both symbolic and nonsymbolic number representations are bound together by associations to verbal number labels. Such a configuration could represent a developmental precursor of the triple-code model of numerical cognition (Dehaene, 1992) within which symbolic forms of number, visual (Arabic) and verbal numerals, operate in interaction with analogue magnitude representations. This proposal is not inconsistent with the account offered by Geary & vanMarle (2018), whereby increasing skill in object counting (using spoken numerals) from age 3;10 to 5;4 promotes acceleration in symbolic comparison skills. However, it is important to note a critical difference between Geary & vanMarle (2018) and the current study. While 64% of four-year-olds in the current sample were able to name nine or ten of Arabic numerals 1-10 (total sample mean score 8.02, SD 2.67), 38% of the sample in Geary & vanMarle (2018), at a similar age, were unable to name two numerals. This important difference may be due to the fact that Geary & vanMarle (2018) sampled children at risk for school failure. Our findings concerning the factor structure of magnitude comparison appear to conflict with those of Matejko & Ansari (2016), who found divergent trajectories across the first year of schooling. However, there is also agreement between the findings. Both studies find high correlations between the tasks, and both conclude that nonsymbolic numerosity comparison is not a predictor of arithmetic, as had been proposed (e.g. Mazzocco et al. 2011). In our study, we attribute the development from a two factor to a one factor model not to overlapping representations, but to a shared framework of magnitude comparison, which depends on the creation of mappings between nonsymbolic magnitude representations and corresponding Arabic numerals. While Matejko & Ansari (2016) suggest that the nature of this learning process is an open question, we propose that the integration of digits, number words and magnitudes plays a critical role (see also Malone, Heron-Delaney, Burgoyne, & Hulme, 2019).

A further implication of our findings is related to the recent study by Yuan et al. (2019), which offers a more detailed understanding of the developmental relations between physical quantities, spoken numbers and written numbers (the elements of the triple code model). Yuan et al. (2019) argue that creating associations between single-digit Arabic numerals, spoken words and quantity representations in an ANS are critical for numeracy development (see also Malone et al., 2019). However, they provide strong evidence against the idea that forming associations between quantities and symbols forms the basis for progressing to use multi-digit numbers. Our findings are consistent with this position insofar as, in our data, pre-school knowledge of the association between spoken and written forms (including multi-digit numbers) is a strong predictor of later arithmetic skills, over and above any prediction by tasks tapping participants' processing of symbolic or non-symbolic magnitudes. Our findings and those of Yuan et al. (2019) suggest that early formal arithmetic skills draw on relatively advanced knowledge of Arabic numeral structure, and that cues to this structure are provided by the spoken forms of multi-digit numbers, not by representations of their magnitudes.

We acknowledge a number of limitations in the current study. We have no measure of object counting, which has recently been shown to be an influential preschool predictor of later arithmetic (Chu et al., 2018). However, the levels of number processing recorded in our sample would suggest that ceiling effects would most likely have been found in object counting. We also acknowledge that our study includes only a limited range of domain-general factors, and did not include measures of working memory and attention. We further acknowledge that we have focussed on arithmetic skills as our sole outcome, at the expense of other important areas of mathematical development which merit consideration (LeFevre et al. 2010). It is also the case that the size of our sample is relatively small, and that our findings would be substantially strengthened by replication using a larger study.

In summary, we have assessed the development of numerical cognition between preschool (age 4) and the end of the first year of formal schooling (age 6). During this critical developmental period we found a change in the structure of magnitude comparison skills from a two factor model, with independent symbolic and nonsymbolic components, to a unitary structure, consistent with the mapping of Arabic numeral representations onto nonsymbolic representations of numerosity. Concerning the longitudinal prediction of arithmetic skills we found that a single latent factor, Arabic numeral knowledge, reflecting the shared variance from numeral writing, reading and identification (including multi-digit numerals), was the sole predictor of arithmetic outcome. We propose that learning the associations between spoken and Arabic numerals (which necessitates understanding of the complex structural correspondences between spoken and Arabic multi-digit numerals) provides one of the critical foundations for arithmetic development in young children. Assessment of pre-schoolers' Arabic numeral knowledge (which can be done quickly with high reliability) is a very powerful predictor of children's later formal arithmetic skills.

References

Bishop, D. (2003). *Test for Reception of Grammar-Version 2 (TROG-2).* San Antonio, TX: The Psychological Corporation.

Brigstocke, S., Moll, K., and Hulme, C. (2016). *TOBANS: Test of Basic Arithmetic and Numeracy Skills*. Oxford, UK: Oxford University Press.

Bull, R., and Scerif, G. (2001). Executive functioning as a predictor of children's mathematical ability: Inhibition, switching, and working memory. *Developmental Neuropsychology, 19,* 273-293.

Carey S. (2009) Where Our Number Concepts Come From. *The Journal of Philosophy*. 106: 220-254.

Chu, F.,vanMarle, K., Rouder, J. & Geary, D.. (2018). Children's early understanding of number predicts their later problem-solving sophistication in addition. *Journal of Experimental Child Psychology*. 169. 73-92.

Caravolas, M., Lervåg, M., Mousikou, P., Efrim, C., Litavský, M., Onochie-Quintanilla, E., Salas, N., Schöffelová, M., Defior, S., Mikulajová, M., Seidlová-Málková, G., and Hulme, C. (2012). Common Patterns of Prediction of Literacy Development in Different Alphabetic Orthographies. *Psychological Science, 23,* 678-686.

Cirino, P. (2011)The interrelationships of mathematical precursors in kindergarten. *Journal of Experimental Child Psychology.* 108, 713-733

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition*, *162*, 12–26. https://doi.org/10.1016/j.cognition.2017.01.014

Dal Martello, M. F. (1990). The dictation of italian numerals. *Language and Cognitive*

Processes, *5*(3), 237–254. https://doi.org/10.1080/01690969008402106

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*. 44: 1-42

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

De Smedt, B., Noël, M. P., Gilmore, D., and Ansari, D. (2013). How do symbolic and nonsymbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2,* 48–55

Donlan, C., Bishop, D. V. M., and Hitch, G. J. (1998). Magnitude comparison by children with specific language impairments: Evidence of unimpaired symbolic processing. *Journal of Language and Communication Disorders, 33*, 149-160.

Donlan, C., Cowan, R. Newton, E. J., and Lloyd, D. (2007). The role of mathematical development: Evidence from children with specific language impairments. *Cognition, 103*, 23-33.

Dunn, L. M., Dunn, D. M., and Styles, B. (2010). *British Picture Vocabulary Scale (3rd ed.).* London, England: GL Assessment.

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, *123*(1), 53–72. https://doi.org/10.1016/j.jecp.2014.01.013

Geary, D. C., & vanMarle, K. (2018). Growth of symbolic number knowledge accelerates after children understand cardinality. *Cognition*, *177*(April), 69–78. https://doi.org/10.1016/j.cognition.2018.04.002

Göbel, S. M., Watson, S. E., Lervåg, A., & Hulme, C. (2014). Children's arithmetic development: it is number knowledge, not the approximate number sense, that counts. *Psychological Science*, *25*(3), 789–798. https://doi.org/10.1177/0956797613516471

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*(1), 17–29. https://doi.org/10.1016/j.jecp.2008.04.001

Hirsch, S., Lambert, K., Coppens, K., & Moeller, K. (2018). Basic numerical competences in large-scale assessment data: Structure and long-term relevance. *Journal of Experimental Child Psychology*, *167*, 32–48. https://doi.org/10.1016/j.jecp.2017.09.015

Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling,* New York, NY: Guilford Press.

Koponen, T., Aunola, K., Ahonen, T., & Nurmi, J.-E. (2007). Cognitive predictors of single-digit and procedural calculation skills and their covariation with reading skill. *Journal of Experimental Child Psychology*, *97*(3), 220–241. https://doi.org/10.1016/j.jecp.2007.03.001

Koponen, T., Salmi, P., & Eklund, K. (2013). Counting and RAN: Predictors of Arithmetic Calculation and Reading Fluency. *Journal of Educational Psychology, 105*(1), 162-175. https://doi.org/10.1037/a0029285

Krajewski, K., & Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learning and Instruction*, *19*(6), 513–526. https://doi.org/10.1016/j.learninstruc.2008.10.002

LeFevre, J.-A., Fast, L., Skwarchuk, S.-L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: longitudinal predictors of performance. *Child Development*, *81*(6), 1753–1767. https://doi.org/10.1111/j.1467-8624.2010.01508.x

Le Corre M, Carey S. One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition*. 105: 395-438.

Lichtenberg, E. & Smith, R. (2005). *Essentials of WIAT-11 and KTEA-11 assessment*. Hoboken:John Wiley.

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*(5), 714–726 https://doi.org/10.1111/desc.12152

Malone, S. A., Heron-Delaney, M., Burgoyne, K., & Hulme, C. (2019). Learning correspondences between magnitudes, symbols and words: Evidence for a triple code model of arithmetic development. *Cognition, 187, 1-9.* https://doi.org/10.1016/j.cognition.2018.11.016

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Impaired Acuity of the Approximate Number System Underlies Mathematical Learning Disability (Dyscalculia). *Child Development*, *82*(4), 1224–1237. https://doi.org/10.1111/j.1467-8624.2011.01608.x

Matejko, A. A., & Ansari, D. (2016). Trajectories of symbolic and nonsymbolic magnitude processing in the first year of formal schooling. *PLoS ONE*, *11*(3), 1–15.

https://doi.org/10.1371/journal.pone.0149863

Miller, K. F., & Stigler, J. W. (1987). Counting in Chinese: cultural variation in a basic skill. *Cognitive Development*, *2*, 279–305.

Moyer, R. S., and Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature, 215,* 1519-1520.

Miura, I.T. & Okamoto, Y. (1989). Comparison of U.S. and Japanese first graders' cognitive representation of number and understanding of place value. *Journal of Educational Psychology*, 81, 109-113 .

Mix, K. S., Prather, R. W., Smith, L. B., & Stockton, J. D. (2014). Young Children's Interpretation of Multidigit Number Names: From Emerging Competence to Mastery. *Child Development*, *85*(3), 1306–1319. https://doi.org/10.1111/cdev.12197

Muthén, L.K. and Muthén, B.O. (1998-2015). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén

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

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences*, *14*(12), 542–551. https://doi.org/10.1016/j.tics.2010.09.008

Raven, J. C, Court, J. H and Raven, J. C. (1993). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Oxford: Oxford Psychologists Press.

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of nonsymbolic and symbolicnumerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science, 20*(3), e12372

Sekuler, R., and Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development, 48,* 630-633.

Simanowski, S., & Krajewski, K. (2017). Specific Preschool Executive Functions Predict Unique Aspects of Mathematics Development: A 3-Year Longitudinal Study. *Child Development*, *85,* 544-561. https://doi.org/10.1111/cdev.12909

UK Government Statistics (2015) *Indices of deprivation*. Retrieved from https://www.gov.uk/government/statistics/english-indics-of-deprivation-2015

Vanbinst, K., Ansari, D., Ghesquière, P., & De Smedt, B. (2016). Symbolic Numerical Magnitude Processing Is as Important to Arithmetic as Phonological Awareness Is to Reading. *PloS One*, *11*(3), e0151045. https://doi.org/10.1371/journal.pone.0151045

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

Wynn, K. (1990). Children's understanding of counting. *Cognition*, *36*, 155–193.

Xenidou-Dervou, I., Molenaar, D., Ansari, D., van der Schoot, M., & van Lieshout, E. C. D. M. (2017). Nonsymbolic and symbolic magnitude comparison skills as longitudinal predictors of mathematical achievement. *Learning and Instruction*, *50*(Supplement C), 1–13. https://doi.org/10.1016/j.learninstruc.2016.11.001

Zhang, X., Koponen, T., Räsänen, P., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2014). Linguistic and spatial skills predict early arithmetic development via counting sequence knowledge. *Child Development, 85,* 1091-1107

Xu, F., & Spelke, E. S. (2000). Large number comparison in human 6-month-old. *Cognition*, *74*, B1–B11.

Appendix: Stimuli for Number Identification Task.

9	6	8	3
82	208	20	28
206	260	26	2060
706	17	7	70
19	119	91	9
12	22	1	2
41	42	14	4
502	25	52	5
1	101	111	11
96	69	6	49
37	13	713	73
7800	807	870	78
43	4	34	304
17	174	74	7
1068	618	18	168
13	3	30	33
58	850	5	85

Table 1

Mean, standard deviations and multiple comparison tests for all Time 1 measures for participants who took part only at Time 1 (T1 only) , compared to those who took part in the full Longitudinal Study (LS).

		T1 only group n=29 (14 male)	LS group n=71 (34 male)	Multiple Comparisons Bonferroni corrected critical value .004 (for overall alpha .05)
		M (SD)	*M (SD)*	
Age in months at T1		50.61 (3.67)	50.59 (3.49)	t (98) = .015 p = .988 d = .006
Nonverbal IQ	Raven's CPM	6.41 (1.70)	6.46 (1.52)	t (98) = .147 p = .883 d = .03
Language Comprehension	TROG-2	3.31 (2.45)	3.08 (2.72)	t (98) = .387 p = .699 d = .09
Vocabulary	BPVS-III	62.55 (17.64)	56.51 (16.20)	t (98) = 1.65 p = .102 d = .36
Arabic Numeral Knowledge	Numeral Writing	7.09 (5.46)	6.76 (5.95)	t (98) = .254 p = .806 d = .06

	Numeral Reading	7.59 (3.26)	8.20 (2.39)	t (98) = .130 p = .897 d = .21
	Numeral ID	7.21 (2.32)	7.28 (2.72)	t (98) = 1.04 p = .301 d = .03
Counting	Rote Counting	13.21 (6.08)	15.42 (14.73)	t (98) = .781 p = .436 d = .20
Symbolic Numeral Comparison	Digit Close	9.83 (3.20)	10.23 (3.23)	t (98) = .561 p = .576 d = .12
	Digit Far	10.52 (3.57)	10.82 (3.62)	t (98) = .377 p = .707 d = .08
	NS FS Close	10.93 (2.17)	9.99 (2.11)	t (98) = 2.02 p = .046 d = .44
Nonsymbolic Numerosity Comparison	NS FS Far	13.21 (2.16)	12.73 (2.72)	t (98) = .837 p = .405 d = .19
	NS SA 2:3	11.69 (2.63)	11.31 (2.30)	t (98) = .710 p = .479 d = .15
	NS SA 3:4	11.24 (2.10)	10.63 (1.10)	t (98) = 1.36 p = .176 d = .36

Notes. M = mean age. SD = standard deviation. All scores are presented as raw scores. For the Magnitude Comparison Tasks: NS = nonsymbolic. FS = fixed size trials. SA = surface-area matched trials.

Table 2

Mean and standard deviations of predictor and criterion measures at all time points

		Time 1	Time 2	Time 3
		n = 100	n = 75	n = 71
		M (SD)	*M (SD)*	*M (SD)*
Nonverbal IQ	Raven's CPM	6.45 (1.57)		
Language Comprehension	TROG-2	3.15 (2.63)*		
Vocabulary	BPVS-III	58.26 (16.77)*		
Arabic Numeral Knowledge	Numeral Writing	6.86 (5.79)*		
	Numeral Reading	8.02 (2.67) [45]*		
	Numeral ID	7.26 (2.60)*		

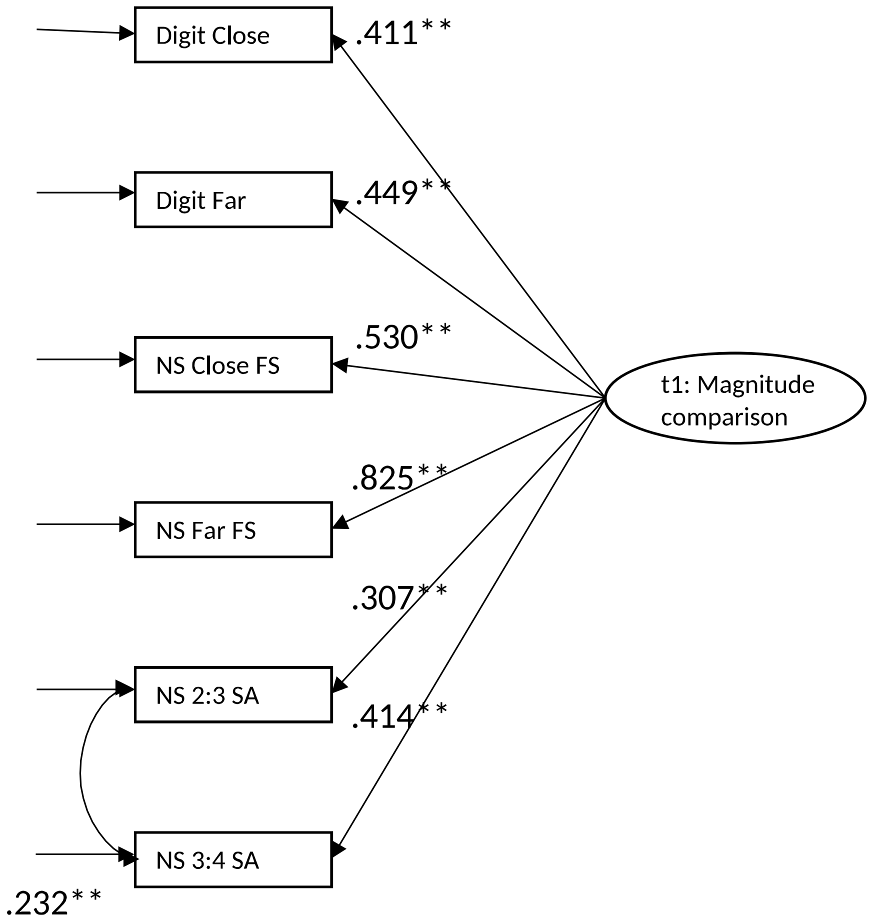
	Rote Counting	14.78 (12.84)*		
Symbolic Numeral Comparison	Digit Close	10.11 (3.21) [7]*		15.61 (4.66)
	Digit Far	10.73 (3.59) [17]*		20.44 (6.39) [2]
Nonsymbolic Numerosity Comparison	NS FS Close	10.26 (2.16) [1]*		13.58 (4.57)
	NS FS Far	12.87 (2.57) [17]*		20.58 (5.98) [2]
	NS SA 2:3	11.42 (2.42) [5]*		17.55 (6.03)
	NS SA 3:4	10.81 (2.03) [2]*		15.58 (5.79)
	Addition Tasks		A: 5.20 (2.55) [5]*	
			B: 5.00 (2.60) [2]*	
	TOBANS			
	Addition			10.41 (6.96)
	Addition w/ carry			4.04 (3.96)
	Subtraction			7.30 (4.40)
	WIAT			4.07 (2.15)

*Notes. M = mean age. SD = standard deviation * individually administered tasks. The number of children scoring at ceiling are shown in square brackets. All scores are presented as raw scores. For the Magnitude Comparison Tasks: NS = nonsymbolic. FS = fixed size trials. SA = surface-area matched trials. All Magnitude Comparison tasks at T1 have 16 trials. All Magnitude Comparison tasks at T3 have 36 trials.*

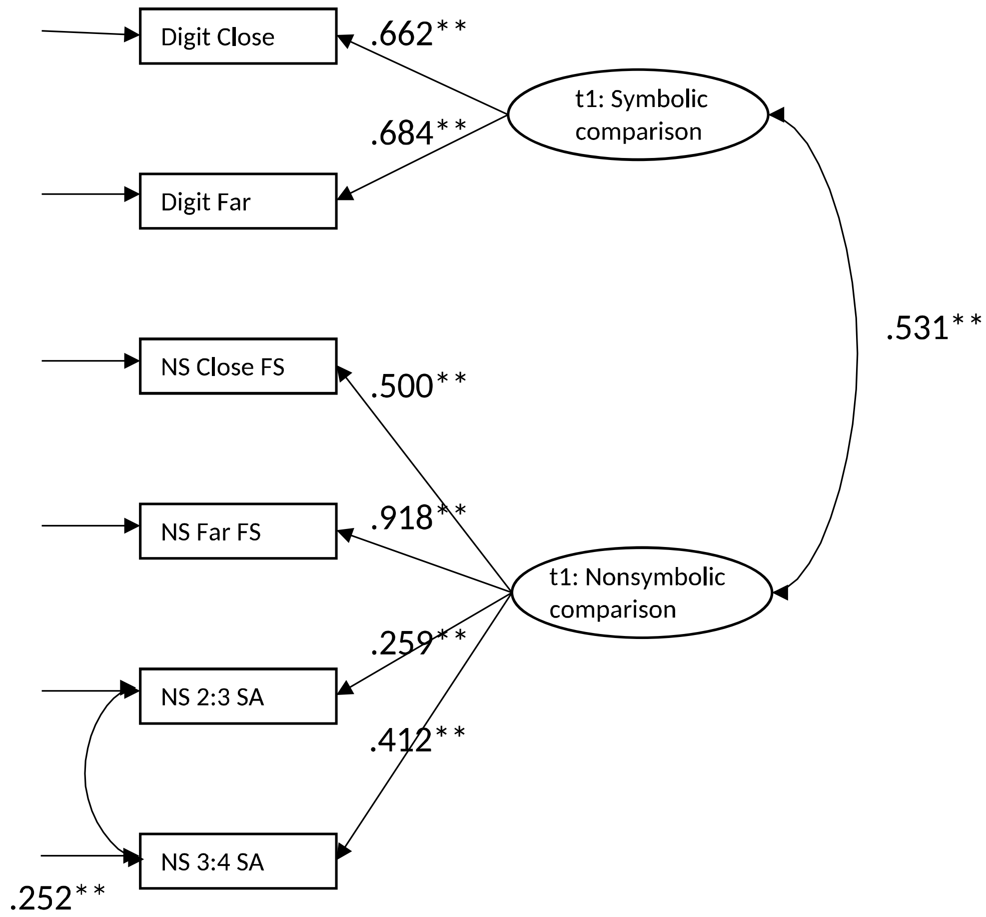
Table 3: *Correlations between the latent variables in Figure 3*

	1	2	3	4	5	6	7	8
1. Nonverbal IQ	---	.228	.392**	.185	.249	.256*	.269	.252
2. Language		----	.507**	.286	.619**	.442**	.447**	.332**
3. Arabic Numeral Knowledge			----	.657**	.597**	.388**	.673**	.641**
4. Counting				----	.390**	.242	.441**	.421**
5. Symbolic Numeral Comparison					----	.538**	.617**	.398**
6. Nonsymbolic Numerosity Comparison						----	.364**	.256**
7. Arithmetic T2							----	.470**
8. Arithmetic T3								----

Notes. Pearson product-moment correlation coefficient. All variables entered are raw scores. * $p < .05$. ** $p < .01$



Digit Close
Digit Far
NS Close FS
NS Far FS
NS 2:3 SA
NS 3:4 SA
t1: Magnitude comparison
.411**
.449**
.530**
.825**
.307**
.414**
.232**



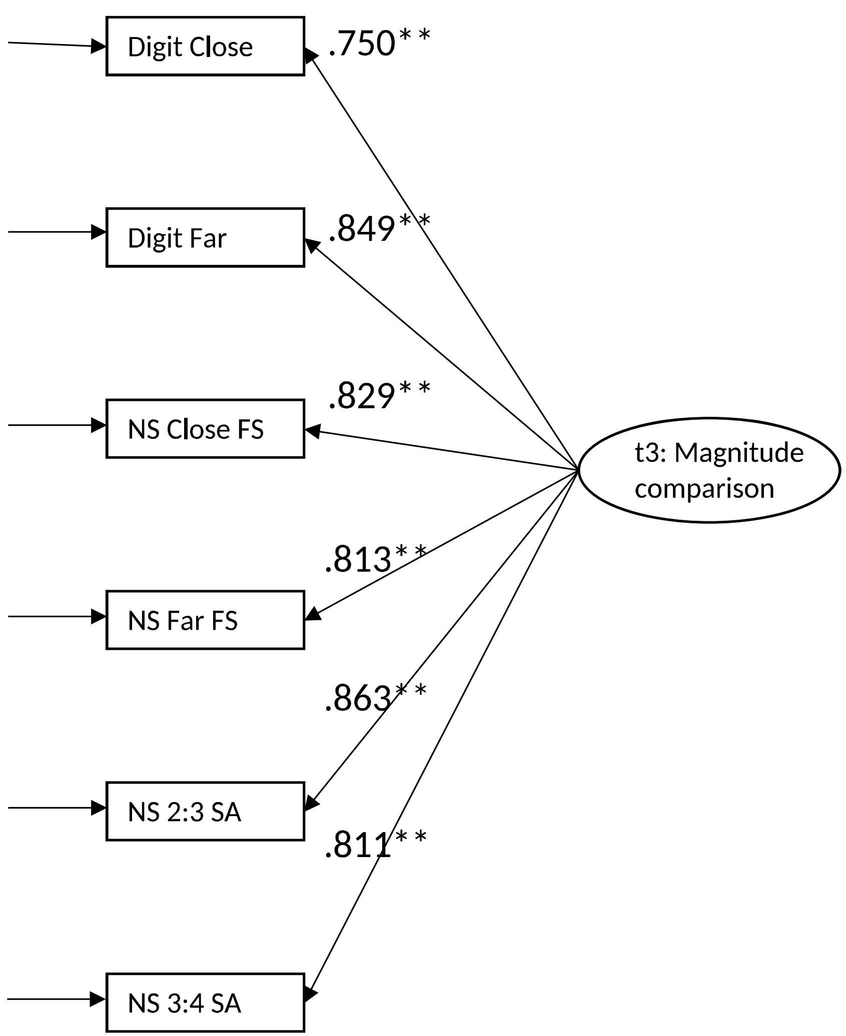
Digit Close
Digit Far
NS Close FS
NS Far FS
NS 2:3 SA
NS 3:4 SA
t1: Symbolic comparison
t1: Nonsymbolic comparison
.662**
.684**
.500**
.918**
.259**
.412**
.252**
.531**

Figure 1. One factor (left side) and two factor (right side) CFA of magnitude comparison tasks (Time 1).

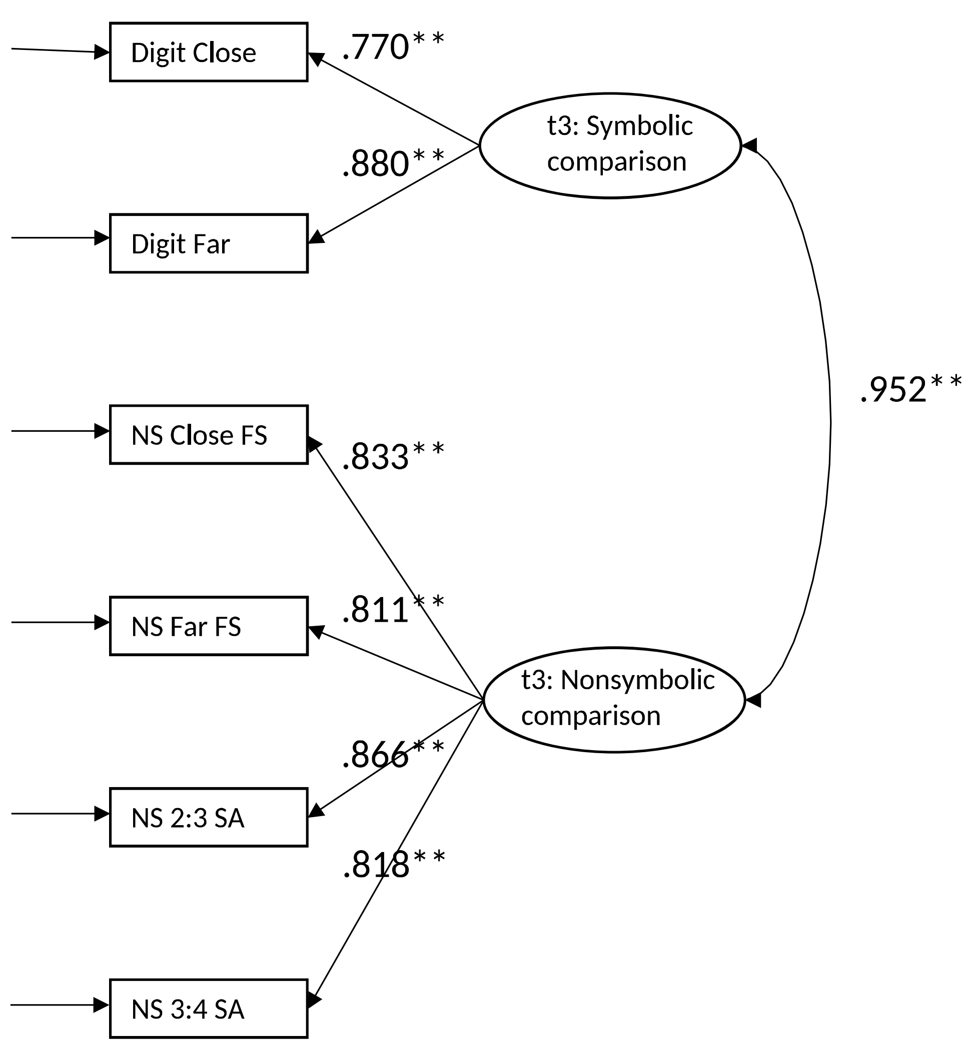
One-factor model fit: $RMSEA$ = .136 (90% CI = .072 - .203), CFI = .827, $SRMR$ = .071.

Two-factor model fit: $RMSEA$ = .058 (90% CI = .000 - .144), CFI = .973, $SRMR$ = .047.

* $p < .05$, ** $p < .01$



Digit Close
Digit Far
NS Close FS
NS Far FS
NS 2:3 SA
NS 3:4 SA
.750**
.849**
.829**
.813**
.863**
.811**
t3: Magnitude comparison



Digit Close
Digit Far
NS Close FS
NS Far FS
NS 2:3 SA
NS 3:4 SA
.770**
.880**
.833**
.811**
.866**
.818**
t3: Symbolic comparison
t3: Nonsymbolic comparison
.952**

Figure 2. One factor (left side) and two factor (right side) CFA of magnitude comparison tasks (Time 3)

One-factor model fit: *RMSEA = .064 (90% CI = .000 - .132), CFI = .991, SRMR = .023*.

Two-factor model fit: *RMSEA = .056 (90% CI = .000 - .130), CFI = .994, SRMR = .021*.

* $p < .05$, ** $p < .001$.

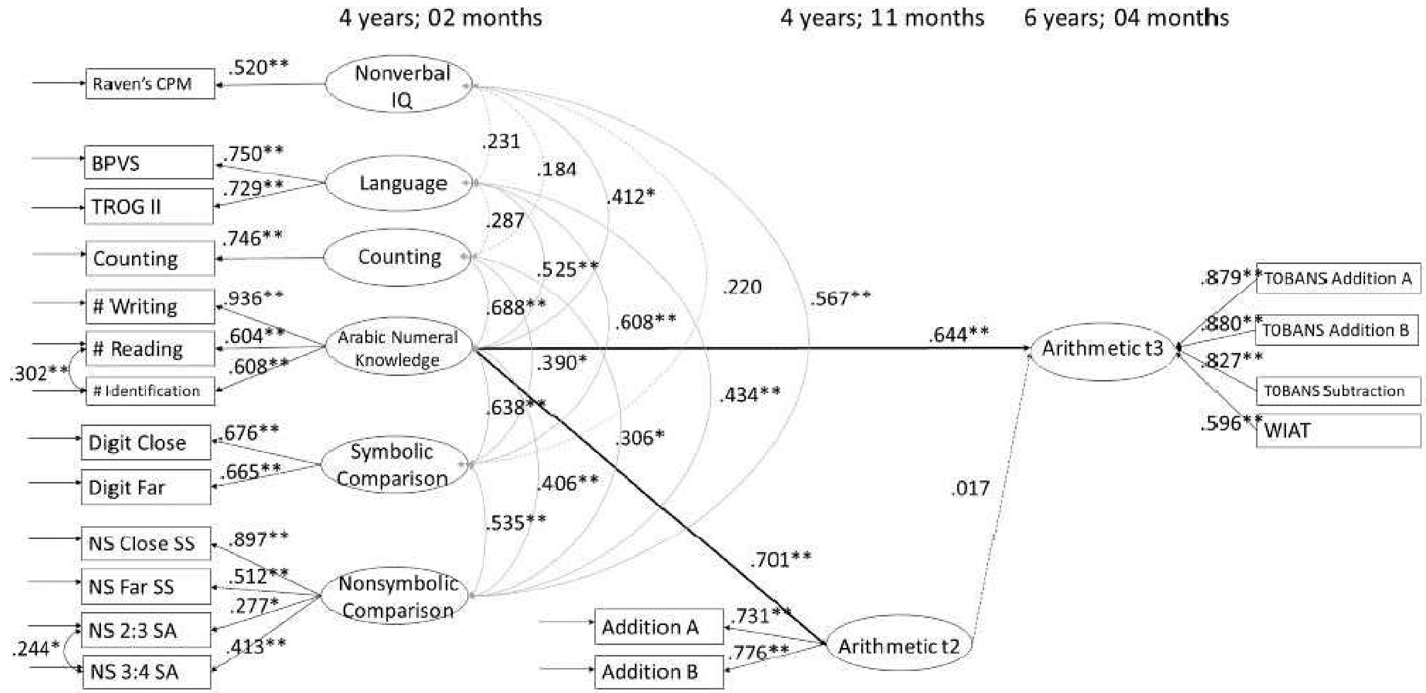

4 years; 02 months
4 years; 11 months
6 years; 04 months
Raven's CPM
.520**
Nonverbal IQ
BPVS
TROG II
.750**
.729**
Language
Counting
.746**
Counting
Writing
Reading
Identification
.302**
.936**
.604**
.608**
Arabic Numeral Knowledge
Digit Close
Digit Far
.676**
.665**
Symbolic Comparison
NS Close SS
NS Far SS
NS 2:3 SA
NS 3:4 SA
.244*
.897**
.512**
.277*
.413**
Nonsymbolic Comparison
.231
.184
.412*
.287
.525**
.220
.567**
.688**
.608**
.390*
.638**
.434**
.306*
.406**
.535**
.644**
.701**
.017
Arithmetic t3
.879**
.880**
.827**
.596**
TOBANS Addition A
TOBANS Addition B
TOBANS Subtraction
WIAT
Arithmetic t2
Addition A
Addition B
.731**
.776**

Fig. 3. *Prediction of arithmetic skills at Time 2 (T2) and Time 3 (T3) by all constructs measured at Time 1 (T1). Latent variables are displayed as ellipses and observed variables as rectangles. Loadings onto hypothesised latent variables are depicted by one-headed arrows from the latent to the observed variables and residuals of each construct by the one-headed arrow into the latent variable. One-headed arrows between latent variables reflect the association and the way of the arrow shows the way of the regression. Two-headed arrows reflect correlation between constructs. Correlations between predictors are shown by grey lines. Solid lines illustrate statistically significant relationships, and dashed lines illustrate statistically nonsignificant relationships. For magnitude comparison, performance was measured on symbolic tasks in which difference was close (digits close) and far (digits far), and nonsymbolic tasks included fixed-size close in number (NS Close FS) and far (NS Far FS), and different ratios of squares with surface-area matched (NS 2:3 SA and NS 3:4 SA). Asterisks indicate significant paths (*$p < .05$, **$p < .01$). BPVS = British Picture Vocabulary Scale, third edition (Dunn, Dunn, & Styles, 2009); Raven's CPM = Raven's Coloured Progressive Matrices (Raven, Raven, & Court, 1998); WIAT = Wechsler Individual Achievement Test, second U.K. edition (Wechsler, 2005)*