Domain-general and domain-specific influences on emerging numerical cognition: Contrasting uni- and bidirectional prediction models

Coolen, I.[1+], Merkley, R.[2+], Ansari, D.[3], Dove, E.[1], Dowker, A.[1], Mills, A.[1], Murphy, V.[4], von Spreckelsen, M.[1], & Scerif, G.[1]

[In press – *Cognition*]

[+]Joint first authors

Affiliations:

[1]Department of Experimental Psychology, University of Oxford, United Kingdom

[2]Institute of Cognitive Science, Carleton University, Canada

[3]Department of Psychology & Faculty of Education, The University of Western Ontario, Canada

[4]Oxford Department of Education, University of Oxford, United Kingdom

Acknowledgements

We would like to thank all the pre-school practitioners, children and parents for their time, enthusiasm and participation in this project, without whom this work would not have been possible. We particularly thank them because our study was disrupted mid-flow by closure of our place of work: an excellent research team and support from nursery settings made testing possible… without a building. Dr Steven Howard was central in our adoption of measures from the Early Years Toolbox and advice on how to use them, so we are extremely grateful to him. We are also very grateful to the four great internship students who helped with data collection: Rebecca Ashworth, Freya Prentice, Xanthippi Vassiliou and Ruxandra Vasile. We also thank the Nuffield Foundation for funding this project through grant EDO 42717 to GS.

The Nuffield Foundation funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. More information is available at: www.nuffieldfoundation.org

Corresponding Author:

Gaia Scerif, Attention, Brain and Cognitive Development Group

Department of Experimental Psychology, University of Oxford

Anna Watts Building, OX2 6GG, Oxford, UK

Phone: +44 [0] 1865 271403

Lab: www.psy.ox.ac.uk/abcd

CRediT Statement: Ilse Coolen: Data Curation, Formal Analysis, Writing; Rebecca Merkley: Conceptualization, Investigation, Writing; Daniel Ansari: Conceptualization, Writing; Emma Dove: Data Acquisition, Supervision; Ann Dowker: Conceptualization, Investigation, Writing; Annelot Mills: Investigation, Data Acquisition, Writing, Murphy: Conceptualization, Investigation, Writing; Megan von Spreckelsen: Data Acquisition, Supervision; Gaia Scerif: Conceptualization, Funding Acquisition; Investigation, Project Administration, Writing.

Abstract

Domain-general skills such as executive functions (EFs), and domain-specific skills such as non-symbolic number sense and symbolic understanding are often pitted against each other as predictors of emerging maths. Here we aimed to investigate early childhood relations between these foundational skills with a balanced, longitudinal design. One hundred and seventy 3- and 4-year-old-children were tested at two time points, 5 months apart, on four domain-general executive and five domain-specific numeracy tasks. A latent EF factor was a strong predictor of symbolic maths and of their growth. In addition, stronger symbolic maths at Time 1 was correlated with later stronger EF, but symbolic maths did not predict EF growth. Our findings provide novel insights into dynamic interplay between general and specific cognitive skills contributing to preschool maths.

Domain-general and domain-specific influences on emerging numerical cognition: Contrasting uni- and bidirectional prediction models

Individual differences in maths skills can be identified from infancy and preschool (e.g., Dowker, 2005, 2008) and have been shown to predict later academic maths achievement (Duncan et al., 2007). Equally, preschool domain-general skills, such as executive functions (EFs) and attention, have also been shown to predict later maths achievement and seem to have a crucial role in the development of maths skills (e.g., Mulder, Verhagen, Van der Ven, Slot, & Leseman, 2017). However, domain-general and domain-specific predictors of maths skills have often been studied in isolation of one another. It is not fully understood how these early maths skills (i.e., domain-specific skills) and domain-general skills interplay to contribute to maths learning in preschool. The main aim of this study was to gain insight into the interactions between foundational domain-general and domain-specific cognitive skills in preschool children. This was done by assessing various domain-general (e.g., inhibition skills, attention and working memory) and domain-specific skills (e.g., cardinal number knowledge, counting, digit identification) that are known to be strong predictors of math achievement longitudinally (Lau et al., in press; Merkley & Ansari, 2016). A longitudinal design allowed us to investigate the factor structure of domain-general and domain-specific skills in order to better understand the development of these inter-related foundational skills.

A number of prior studies exploring domain-general and domain-specific skills longitudinally lack the consistency of measurements across the different time points by, for example, investigating *either* domain-general *or* domain-specific skills predictors as predictors of maths achievement measures in depth, but not both in a fully balanced design (e.g., Chu, vanMarle, & Geary, 2015; Clark, Sheffield, Wiebe, & Espy, 2013; Geary, 2011; Gimbert, Camos, Gentaz, & Mazens, 2019). By using the same measurements at two time points, and by giving equivalent depth to both domain-general and domain-specific skills, the current study allowed us to investigate the directionality of the interplay (i.e., unidirectional or bidirectional predictive relations) between foundational domain-general and domain-specific skills longitudinally.

It makes intuitive sense that domain-general cognitive skills such as organising and following a planned sequence of actions, selecting relevant material and characteristics for attention and ignoring others, and flexibility of problem-solving would be important for maths achievement, as multiple studies have now shown (Blair & Razza, 2007; Bull et al., 2011; Bull & Scerif, 2001; Clark et al., 2013; Fuhs et al., 2014; Schmitt et al., 2017; Welsh, Nix, Blair, Bierman, & Nelson, 2010; Purpura, Schmitt, & Ganley, 2017). It may be less intuitively obvious that there might be a relation in the other direction: that numerical abilities such as counting, number knowledge and arithmetic might predict the development of domain-general cognitive skills. And yet several recent studies do suggest bidirectional longitudinal relations between individual differences in maths achievement measures and EFs measures in preschool and primary school children (Clements et al., 2016; Fuhs et al., 2014; Hassinger-Das, Jordan, Glutting, Irwin, & Dyson, 2014; McKinnon & Blair, 2019; Prager, Sera, & Carlson, 2016; Schmitt, Geldhof, Purpura, Duncan, & McClelland, 2017; Welsh et al., 2010). This suggests that acquiring and practicing mathematics skills may also help refine the control of attention and other EFs. It has been hypothesized that domain-general cognitive skills and domain-specific mathematics knowledge dynamically interact as children progress in their maths learning (Merkley, Matusz, & Scerif, 2018). Specifically, domain-general control processes allow children to select information relevant for learning, but this accumulating knowledge also guides how attentional resources are allocated to task-relevant stimuli and therefore relates to academic achievement (e.g., Amso & Scerif, 2015; Shing & Brod, 2016). Indeed, bidirectional relations between academic achievement and cognitive skills have recently been supported with the mutualism model (Peng & Kievit, in press), and, similarly, the transactional model (Miller-Cotto & Byrnes, 2019). In both the mutualism and transactional models, domain-general cognitive skills and academic abilities influence each other and support growth reciprocally over development. For example, Kievit and colleagues (2019) found that there was a reciprocal relation between 6- to-8-year-old children's development of reasoning abilities and their vocabulary size. Miller-Cotto and Byrnes (2019) also found strong evidence for bidirectional relations between working memory indices and maths / reading achievement measures in very large samples of US-based children from the age of 5 years. These findings lend further support to the hypothesis that cognitive development arises through dynamic interplay between domain-general and domain-specific cognitive skills. We predicted that we would replicate this interplay within this study focusing on executive functions and early maths skills. Of interest, bidirectional relations are reported to strengthen with age (Peng & Kievit, 2020). They have also not been investigated in very young children outside the context of academic achievement or multi-componential assessment measures that are appropriate for what children of a particular age are expected to know and may be variable over age. These are likely to index a mixture of overall knowledge learned in an educational setting, as well as speeded performance factors, in addition to the ability to understand specific numerical concepts. The aim of the current study was therefore to investigate the relationships between foundational maths and domain-general cognitive skills in early childhood, before the start of formal schooling, and test whether a bidirectional model is a better fit than a unidirectional one.

The study hypotheses were three-fold. Firstly, we hypothesised that different domain-general (i.e., EFs and attention, for brevity "executive skills") and domain-specific cognitive skills (i.e., cardinality, symbolic number knowledge, non-symbolic numerical skills and counting, for brevity "maths skills") load respectively onto two separable factors, providing a good fit for the data. Importantly, the structure of EFs in early childhood (~3- to-5-year-olds) has been found to be best fit by a unitary factor (Hughes & Ensor, 2011; Wiebe et al., 2011). This EF factor is thought to differentiate over development, and EFs seem to be best explained by a three-factor model of separate, but overlapping skills, in middle childhood (~8- to-13-year-olds) (e.g., Hartung et al., in press; Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; but see Lee et al., 2013 for results supporting a two-factor model through middle childhood). Secondly, it was hypothesised that the latent factor indexing domain-general skills at Time 1 would predict latent factor scores indexing domain-specific skills both at Time 1 and at Time 2. Furthermore, we expected that domain-general scores would predict growth in domain-specific skills (analysed by testing the amount of variance in domain-specific skills at Time 2 explained by domain-general skills at Time 1, while also controlling for individual differences in domain-specific skills at Time 1). In line with recent models applied to older children and highlighted above (Kievit et al., 2019; Miller-Cotto & Byrnes, 2019), our third and final hypothesis stated that while domain-general cognitive skills would predict absolute scores and growth in domain-specific skills, this link would be bidirectional, with domain-specific skills at Time 1 also predicting both absolute executive skills scores at Time 2 and growth in domain-general cognitive skills. This final hypothesis attempted to replicate the recent empirical findings (e.g., Schmitt et al., 2017), as well as the possibility that being better able to deal with numerical challenges may generalise to improvements in executive and attentional skills over time.

Method

Participants

This study is part of a wider research project investigating the influence of cognitive processes and the educational environment on preschool maths skills in both a cross-sectional (N = 231) and a longitudinal design (N = 170). For the current longitudinal part of the study, 170 children aged 3 (88 children) to 4 (82 children) years of age (Time 1: M=47.16 months, Range=39-54 months; Time 2: M=51.51 months, Range=44-59 months) were tested at two time points, with on average 5 months between the time points. Children were recruited from 12 different preschool settings across a county within the United Kingdom. Formal education starts during the September following children's 4th birthday (therefore, on average between 4 and 5 years of age in the UK), when children enter "Reception", which is similar to kindergarten in many other countries. Prior to this, many children attend preschool childcare settings, which are not publicly funded, but are inspected and regulated by the UK Office of Standards in Education (Ofsted). Practice in childcare settings is shaped by the Early Years Foundation Stage (EYFS, https://www.gov.uk/early-years-foundation-stage) framework, which sets early learning goals. Of the 12 settings we recruited from this study, 5 were independently run charity nurseries or preschools, 4 were workplace nurseries on sites of employment and 3 were school-based nurseries. All preschool settings were supplied with information and consent sheets for the setting and the parents. Once informed consent was gained from preschool settings and parents, the participating parents were asked to report demographic information (e.g., postcode, ethnicity of their child(ren)) through parent questionnaires. Eighty-two percent of all participating parents returned the parent questionnaire. These questionnaire responses established that, in this longitudinal sample, 12.9% of the participating children were known to speak one or more languages in addition to English, with 20 different languages reported as second language, of which German, Dutch and Polish were the most frequently reported. It should be noted that some of these languages encode number words differently to English (i.e., four and twenty rather than twenty-four). Due to the large variety of additional languages and to retain the local, regional and national representativeness of our sample, all children were included in the analyses. Furthermore, parents reported that 79% of the children were of Caucasian ethnicity. Based on the

children’s post code, the SES deprivation decile was obtained from the English Index of Multiple Deprivation (English Indices of Deprivation, 2015), with a mean deprivation level of 7.62 (range = 2-10, SD = 2.26). The Index of Multiple Deprivation (IMD) represent statistics on the relative deprivation in England. IMD deciles range from 1 to 10, 1 corresponding to the most deprived decile and 10 the least deprived decile. Within the longitudinal sample, 4 children were reported by parents to have a special educational need or disability (SEND). These children were included in the sample as they were judged by the experimenters to understand the tasks and as the inclusion of SEND children reflects the UK preschool population more accurately than to exclude them.

Materials

Executive and attentional measures

These tasks were chosen to fit best with the age range tested in this study, and also to fit the emerging literature on the structure of EF in the preschool period (e.g., Wiebe et al., 2011) and on predictors of EF from infancy and early childhood (Garon et al., 2008; Hendry et al., 2016), rather than only be inspired by the literature on EF in older children, adolescents and adults (e.g., Miyake & Friedman, 2012).

Go/no-go task

This go/no-go task was designed to assess young children’s inhibitory control (Early Years Toolbox; Howard & Melhuish, 2017). In this task, children were presented with either fish or sharks (1500ms per stimulus) on an iPad screen and were instructed to tap the fish (go-trials) and not tap the sharks (no-go-trials). The task consisted of 20 practice trials (1 block of 5 go-trials; 1 block of 5 no-go-trials and 1 block of 8 go-trials and 2 no-go-trials) and 3 blocks of 25 testing trials (20 go-trials and 5 no-go-trials). By presenting go-trials in 80% of all testing trials, a prepotent tendency to respond is created. Therefore, in case of a no-go-trial (20% of all testing trials), the child is required to inhibit their response. Trials with a reaction time faster than 300ms were removed since these reactions are unlikely to be in response to the stimulus. The product of accuracy data from the go-trials (% accurate responses on go stimuli) and accuracy from the no-go trials (% accurate responses on no-go stimuli) provide a combined measure of inhibition skills (Howard & Melhuish, 2017).

Mr Ant task

The Mr. Ant task is an iPad-delivered task that assesses visuo-spatial short-term memory in young children (Early Years Toolbox; Howard & Melhuish, 2017). In each trial, a series of stickers would appear on a cartoon ant for 5 seconds after which a blank screen appeared for 4 seconds. Next, the cartoon ant reappeared and children were asked to tap the locations on the ant's body where they remembered the previously held stickers to be. The number of stickers started from 1 sticker at level one up to 8 stickers at level 8 with 3 trials per level. The task ended when the last trial of level 8 was reached or after 3 failed trials on the same level. Visuo-spatial short-term memory capacity was generated by a point score, with 1 point for each consecutive level with at least 2 of the 3 trials correct and 1/3 of a point for all accurate trials thereafter (Howard & Melhuish, 2017).

Animal Stroop task

The Animal Stroop task is an adaptation of the classic Stroop task (Stroop, 1935), adjusted by Merkley, Thompson and Scerif (2016) to suit young children. It is an inhibitory control measure that assesses the ability to shift attention away and inhibit task-irrelevant dimensions of stimuli. In this task, children were presented with two differently sized animal images (e.g., mouse and elephant) for 5000ms on a computer screen and were instructed to decide which of the two animals was largest in real life. After a series of practice trials geared to gauge children's understanding and practice speeded responding, twenty-four trials were displayed in a random order and correct side of response, congruency (12 congruent and 12 incongruent trials) and animal size were counterbalanced. Overall accuracy was used as the dependent measure for further analysis (following Merkley et al., 2016).

Cancellation task

A cancellation task, created by Steele, Karmiloff-Smith, Cornish and Scerif (2012) was a visual search task designed to assess selective attention and its relations to early literacy and numeracy. In the current version of the task, we asked children to tap all dogs (exemplar search) / animals (categorical search) in a field of distractors (objects) as fast as they could with a stylus pen on a Windows Surface Pro tablet. There were four runs: two exemplar runs and two category runs. The order (i.e. exemplar runs and categorical runs) was counterbalanced. For further analysis we used a computed score of quality of search (Q score) calculated by taking the square of number of correct responses divided by the product of number of targets and total time spent on the task. A high Q score reflects a combination of a high number of cancelled targets, and a high cancellation speed. Q score has been used successfully as a single index of search performance for adult participants (Dalmaijer et al., 2015) and for children (Woods et al., 2015) as young as three years of age (Doherty et al., 2018). This allows one to select a single search performance index, as is most appropriate for factor analysis, our analytical strategy here. The measure used was the average Q score across the two category runs (following Steele et al., 2012).

Early Mathematics Measures

These measures were selected to tap different skills thought to be foundational in preschool math (see Dowker, 2008), rather than math achievement in a single multi-componential standardised measure.

Give-N task

This task measures cardinal number knowledge and has often been used to assess cardinal knowledge (Wynn, 1992). This task reflects children's ability to give the experimenter a set number of items from a larger group. At the first time of testing (Time 1), children were told that Mr Monkey (a monkey hand puppet) was hungry and instructed to hand a number of edible items (between 1 and 8 out of 15 items, e.g., small plastic bananas) to Mr Monkey. To allow for individual growth over time, at the second testing time-point (Time 2) the requested number of items went up to 16 out of 20 (1-8, 10, 11, 14, and 16). Children had two trials per numerical value. The highest number of items that children handed to the hand puppet correctly on any trials administered was taken as index of cardinal knowledge.

Counting amounts task

The counting amounts task was used to assess counting accuracy, which typically develops at the age of 3-4 years old. Children were shown two sets of a number of objects (apples or strawberries; Time 1 = 2,3,5,6,8; Time 2 = 2,3,5,6,8,10,11,14,16,18) and asked to say how many objects there were. Children were instructed that they could touch or move the objects to count if they wanted to. The task was ended when children produced two consecutive incorrect trials. To index counting accuracy, the total number of correct responses across the two sets of objects was used (max 10 at Time 1 and 20 at Time 2).

Counting high task

Similar to the counting amounts task, the counting high task assessed a specific counting skill, in particular how high children were able to count correctly. Children were encouraged to count as high as they could count and were stopped at 100. To index counting score, the highest number reached without making an error in correct verbal sequence recital was used.

Number naming task

The number naming task was created as part of the Numeracy Screener (Lyons et al., 2018) and is used to measure symbolic number identification. In this task children were asked to name symbolic numbers 1-9 as they were pointed to on a page, in non-sequential order (each number appearing twice on the page). The total number of correctly named digits (out of 18) was used as a number identification measure.

Magnitude comparison task

The magnitude comparison task was created to assess number sense acuity. In this task children were required to guess which of two non-symbolic quantities (e.g., arrays of beach balls, arrays of pennies, arrays of pizza slices) was more numerous. This computerised task comprised of 6 practice trials and 48 testing trials with a break in the middle. Ratios between the displayed quantities were either 0.25, 0.50 or 0.75. Continuous perceptual variables were controlled for by altering object size, with 50% congruent trials (i.e., the more numerous quantity corresponds to the visually larger quantity) and 50% incongruent trials (i.e., the more numerous quantity corresponds to the visually smaller quantity) presented in a random sequence. Overall accuracy was chosen to index number sense acuity as this measure was previously shown to be the most reliable measure for magnitude comparisons in children in our target age range (Gilmore et al., 2014).

General Cognitive Ability

In order to control for individual differences in general cognitive abilities (i.e., intelligence and verbal skills) two control measures were also assessed. The picture similarities subtest from the British Ability Scales – 3rd edition (BAS-3) was used to measure cognitive fluency. In this task, children are required to match a card to one of four target cards per trial on the basis of either perceptual or semantic similarities (e.g., matching a card with a stamp on to one with a letter on). To assess receptive vocabulary (i.e., verbal skills) the British Picture Vocabulary Scale – 3rd edition (BPVS-3) was used. In this task children were shown 4 pictures on a page for each trial and were asked to point to the picture that corresponded with a certain word, orally presented by the experimenter. The difficulty of the task gradually progressed and the task would end when the child produced fewer than 4 correct responses in one set of 12.

Design and procedure

This study is part of a wider study that employed a combined cross-sectional and longitudinal design. The data reported here include all children for whom data was collected longitudinally at two time points. During the first time of testing (Time 1) children's domain-general cognitive skills, domain-specific early number skills, and control measures for cognitive fluency and receptive vocabulary were assessed. Children were revisited for the longitudinal sample approximately 5 months later (Time 2) to reassess the same domain-general cognitive skills and domain-specific early number skills. The control measures were only administered once, as part of Time 1. At each testing time, each child went through two testing sessions (set A: animal Stroop, cancellation, counting amounts, magnitude comparison; set B: give-N, go/no-go, counting high, Mr Ant, number naming), with 1 of 4 possible testing sequences within a set. This study received ethical approval from the local Research Ethics institutional review board (approval No. XXX).

The analyses presented here were pre-registered on the Open Science Framework (https://osf.io/ng3xw/?view_only=685b007f6d684fb0bb2a7fac84c15404, anonymised for blind peer review) and will be reported as planned. The pre-registered analyses consisted of three phases: an initial descriptive phase, a comparison of model fits investigating the structure of preschool skills using Confirmatory Factor Analysis (CFA) at Time 1 and Time 2, and finally an investigation of the longitudinal relations between executive and maths skills (using multiple unidirectional linear regressions and cross-lagged panel models to capture the inter-correlations between latent factors at Time 1 and Time 2). Additional analyses (i.e., one-factor model fit for CFA, Exploratory Factor Analysis (EFA), and further investigation of the cross-lagged model) will be reported as further exploratory analyses in the supplementary materials (Tables S1-S3). Data were analysed using Lavaan version 0.6-3 (Rosseel, 2012) in R version 3.5.1.

Results

Pre-registered Analyses

Descriptive statistics

Table 1 displays descriptive statistics for age and every task measured separately for Time 1 and Time 2.

Table 1. *Descriptive statistics per task for Time 1 and Time 2.*

Task	N	M	SD	Mdn	Range	Min	Max
Age T1 (months)	170	47.16	3.54	47	15	39	54
Age T2 (months)	170	51.51	3.56	51	15	44	59
Go/no-go T1 (impulse control score)	159	0.51	0.22	0.53	0.95	0.00	0.95
Go/no-go T2 (impulse control score)	167	0.64	0.20	0.67	0.92	0.08	1
Mr Ant T1 (short-term memory score)	167	1.50	0.76	1.67	3.33	0.00	3.33
Mr Ant T2 (short-term memory score)	169	1.78	0.77	2.00	3.33	0.00	3.33
Animal stroop T1 (% accuracy)	169	74.82	23.50	83.50	100	0.00	100
Animal stroop T2 (% accuracy)	169	83.07	20.54	96.00	70.50	29.50	100
Cancellation T1 (Q-score)	169	0.61	0.19	0.58	0.99	0.20	1.19

Cancellation T2 (Q score)	164	0.73	0.19	0.72	1.17	0.18	1.35
Give-N T1 (highest / 8)	169	4.93	2.44	5	7	1	8
Give-N T2 (highest / 16)	165	7.01	4.10	6	16	0	16
Count high T1 (highest number)	166	16.74	12.36	13	100	0	100
Count high T2 (highest number)	170	19.61	13.63	14	100	3	100
Count amounts T1 (highest / 10)	164	7.15	2.44	8	10	0	10
Count amounts T2 (highest / 20)	168	9.95	3.94	9.50	18	2	20
Number naming T1 (number correctly named / 18)	167	11.52	6.45	14	18	0	18
Number naming T2 (number correctly named / 18)	169	13.71	5.25	16	18	0	18
Magnitude comp. T1 (overall accuracy)	162	0.77	0.16	0.80	0.67	0.33	1
Magnitude comp. T2 (overall accuracy)	168	0.83	0.12	0.86	0.52	0.48	1
BPVS-3 raw	170	56.61	18.17	57.50	89	12	101

BPVS-3 stnd	167	102.98	13.76	103	59	72	131
BAS-3 raw	169	17.95	4.17	18	18	7	25
BAS-3 stnd	169	92.43	11.84	92	79	62	141

What is the structure of latent mathematics and executive skills in preschool?

Confirmatory Factor Analyses

First, confirmatory factor analyses (CFAs) were used to test whether a unitary factor model encompassing all early numerical and cognitive skills, a two-factor model (Executive Functions (EFs) and early maths skills (maths)) or a three-factor model (EFs, symbolic maths skills and non-symbolic maths skills) fit the data best. This approach addressed the fact that the literature on early numeracy is inconsistent on two topics. 1) It has previously been suggested that preschool cognitive skills (including EF and maths) are difficult to disentangle and would fit better in a unitary factor model. 2) Some debate exists on the question whether non-symbolic maths and symbolic maths skills load onto one factor (i.e., early mathematics skills) (e.g., Kolkman, Kroesbergen, & Leseman, 2013). We note that our study design was not optimised to test a separate non-symbolic latent factor because our test battery included a single non-symbolic indicator, magnitude comparison, but the controversy in the broader numerical cognition literature warranted a cautious approach in assuming that magnitude comparison performance would cluster with symbolic maths skills in these young children at any time-point, a point to which we return later via exploratory factor analysis. At a minimum, testing differential model fit indices for a two- and three-factor model would give us some clues on this highly debated issue.

CFA models were fit using Lavaan version 0.6-3 (Rosseel, 2012) in R version 3.5.1. Missing data (< 2%) were inspected for reasons and were handled using a full information maximum likelihood estimation, shown to provide better fit indices than other imputation and estimation methods in CFA (Köse, 2014). Since the assumption of multivariate normality was violated, (Royston, $H = 501.40$, $p < .001$), with most tasks violating the assumption of univariate normality (with the exception of go/no-go at T1 and the cancellation task at T1 & T2), all models were estimated with robust maximum likelihood and latent variables were standardised allowing free estimation of all factor loadings.

The fit indices for the unitary CFA models at both time points show that the unitary model does not fit the data better than the 2- or 3-factor model splitting up domain-general and domain-specific skills (please see supplementary material for a detailed account of the unitary CFA). Fit indices indicated that the 3-factor CFA model (Time 1: AIC = 5341, Robust CFI = 0.930, Robust TLI = 0.899, and Robust RMSEA = 0.073; Time 2: AIC = 5554, Robust CFI = 0.944, Robust TLI = 0.920, and Robust RMSEA = 0.062) with factors EFs, non-symbolic maths and symbolic maths, significantly fitted the data better than the 2-factor CFA model (Time 1: AIC = 5352, Robust CFI = 0.893, Robust TLI = 0.851 and Robust RMSEA = 0.089; Time 2: AIC = 5567, Robust CFI = 0.897, Robust TLI = 0.858 and Robust RMSEA = 0.082) with factors EFs and maths at Time 1, $\chi^{2\text{diff}}(1) = 10.94$, $p < .001$, and at Time 2, $\chi^{2\text{diff}}(1) = 9.11$, $p < .01$. Figure 1 displays a diagram of the 2-factor CFA model and the 3-factor CFA model.

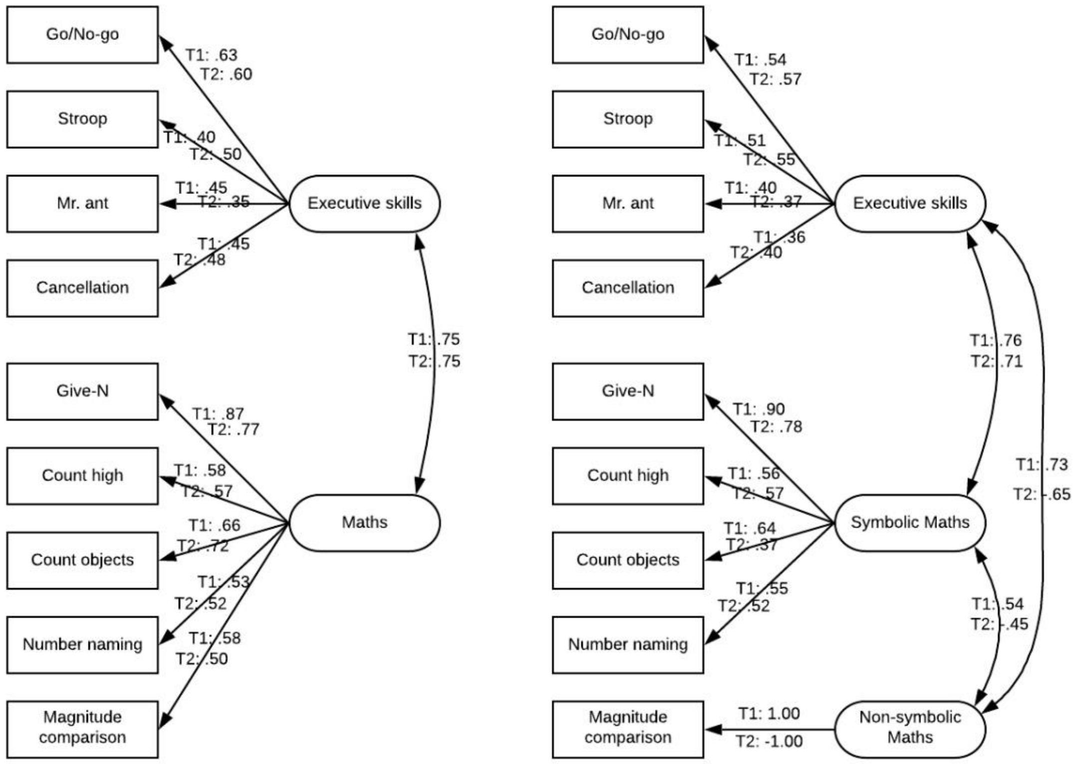

Figure 1. Diagram of the observed variables in the 2-factor model (left) and 3-factor model (right), including the factor loading for each variable for T1 and T2 and the covariance between latent variables.

Given that the 3-factor CFA model best fit the data, factor scores for the latent variables EFs, symbolic maths and non-symbolic maths were extracted to be used in subsequent regression analyses. Table 2 presents standardised factor loadings for each variable at Time 1, as well as the covariance between latent variables and Table 3 displays factor loadings for all variables and covariance between latent variables at Time 2.

Table 2. Factor loadings of 3-factor CFA model for Time 1.

Latent factor	Indicator	B	SE	z-value	β
EFs T1	Go/no-go T1	0.12	0.02	4.22***	0.54
EFs T1	Stroop T1	11.92	1.99	5.33***	0.51

EFs T1	Mr. ant T1	0.31	0.07	3.48***	0.40
EFs T1	Cancellation T1	0.07	0.02	2.40*	0.36
Symbolic maths T1	Give-N T1	2.20	0.17	16.95***	0.90
Symbolic maths T1	Count high T1	6.95	0.96	6.87***	0.56
Symbolic maths T1	Count objects T1	1.57	0.19	8.10***	0.64
Symbolic maths T1	Number naming T1	3.52	0.49	7.40***	0.55
Non-symbolic maths T1	Magnitude comparisons T1	0.16	0.01	20.15***	1.00

Correlations between latent factors

Latent factor	EFs T1	Symbolic maths T1
EFs T1	-	-
Symbolic maths T1	.763***	-
Non-symbolic maths T1	.732***	.537***

Note: *p< .05,**p< .01,***p< .001

Table 3. Factor loadings of 3-factor CFA model for Time 2.

Latent factor	Indicator	B	SE	z-value	β
EFs T2	Go/no-go T2	0.11	0.02	6.07***	0.57
EFs T2	Stroop T2	11.31	1.75	6.50***	0.55
EFs T2	Mr. ant T2	0.28	0.07	3.77***	0.37
EFs T2	Cancellation T2	0.08	0.02	3.06***	0.40
Symbolic maths T2	Give-N T2	3.21	0.31	10.15***	0.78

Symbolic maths T2	Count high T2	7.78	1.08	6.64***	0.57
Symbolic maths T2	Count objects T2	2.86	0.30	9.48***	0.73
Symbolic maths T2	Number naming T2	2.73	0.42	6.28***	0.52
Non-symbolic maths T2	Magnitude comparisons T2	-0.12	0.01	-15.86***	-1.00

Correlations between latent factors

Latent factor	EFs T2	Symbolic maths T2
EFs T2	-	-
Symbolic maths T2	.707***	-
Non-symbolic maths T2	-.651***	-.449***

Note: *p< .05,**p< .01,***p< .001

Although the better fit of the three-factor model clearly highlighted that symbolic skills and magnitude comparison skills are best considered separately, it is problematic to use our non-symbolic measure as a latent variable moving forward, as it derives from a single observed measure. Indeed, the factor loading of the non-symbolic measure at Time 2 was flipped in the analyses and resulted in a negative score ($\beta = -1$). This counterintuitive negative relation is likely to be a distributional artefact of that variable, given that all non-parametric correlations and covariance are positively related (Table S2). Due to these data-driven issues associated with the non-symbolic variable, we could only confidently test our main study question (the directionality of domain-general and domain-specific skills) by focusing on the two latent factors that could be clearly isolated: the EFs latent factor and the Symbolic Maths latent factor.

The non-symbolic measure was not treated as a latent factor and was investigated in more depth in exploratory analyses. An Exploratory Factor Analysis (EFA) indicated that at

Time 1 the magnitude comparison measure loaded both on the EFs factor ($\beta > 0.32$) as well as on the Maths factor ($\beta > 0.42$), although slightly stronger on the Maths factor. However, at Time 2, the magnitude comparison measure only loaded onto the EFs measure ($\beta > 0.54$) (see Unregistered Exploratory Analyses, supplementary materials). Since the Non-symbolic Maths factor was deleted from further analyses, the fit for the 2-factor CFA including the EFs latent factor and the Symbolic Maths latent factor was checked again. Fit indices of the CFA for the two factor-model not including a Non-symbolic Maths latent factor indicated a good model fit (Time 1: AIC = 5537; Robust CFI = 0.949; Robust TLI = 0.924; Robust RMSEA = 0.63; and Time 2: AIC = 5838; Robust CFI = 0.964; Robust TLI = 0.947; Robust RMSEA = 0.51).

What are the longitudinal relations between executive and symbolic mathematics skills?

In the second phase of the analyses, both unidirectional and bidirectional longitudinal relations between executive and symbolic maths skills were assessed using multiple unidirectional linear regressions and a cross-lagged model.

Unidirectional Regressions 1: Executive skills to Symbolic Mathematics?

First, the latent variables, EFs and symbolic maths, were entered in unidirectional hierarchical regression analyses, with EFs T1 predicting maths T2 (model 1a), as well as EFs T1 predicting growth in maths T2 (model 1b), by entering maths T1 as autoregressor.

In the first step, we compared the basic intercept model with a model where symbolic maths at Time 2 was predicted by age. Age at Time 1 was significant as predictor of symbolic maths at Time 2, $F(1, 168) = 7.00$, $p < .01$. In a second step, control measures BAS and BPVS (raw scores), measured at Time 1, were entered and were significant predictors of symbolic maths at Time 2, $F(2, 166) = 23.53$, $p < .001$. Age, BAS and BPVS explained a total of 23.85% of the variance of the model. In step 3 in model 1a, EFs at Time 1 was entered as final step and appeared as a significant predictor of symbolic maths at Time 2 ($\beta = .697$), $F(1, 165) = 87.91$, $p < .001$, improving the model to explain 50.02% of the variance.

In step 3 in model 1b, we tested whether the executive latent factor predicted growth, rather than just outcomes, in symbolic maths. To do so, symbolic maths at Time 1 was entered as autoregressor and improved the model fit significantly ($\beta = .513$), $F(1, 165) = 121.2$, $p < .001$, with 55.83% of the variance explained by the model with age, BAS, BPVS and symbolic maths at Time 1 as predictor of symbolic maths at Time 2. As the fourth and final step to model 1b, EFs at Time 1 was entered, which significantly improved the model fit to explain 56.67% of the variance of the model ($\beta = .233$), $F(1, 164) = 4.20$, $p < .05$.

Unidirectional Regressions 2: Symbolic Mathematics to Executive skills?

Another hierarchical regression was conducted to examine whether symbolic maths at T1 was a predictor of EFs at T2 (model 2a) and whether symbolic maths at T1 was a predictor of growth in EFs at T2 (model 2b), with EFs at T1 as autoregressor.

In step one, we compared the basic intercept model with a model where EFs at Time 2 was predicted by age. Age at Time 1 was significant as predictor of EFs at Time 2, $F(1, 168) = 17.46$, $p < .001$. In a second step, control measures BAS and BPVS (raw scores), measured at Time 1, were entered and were significant predictors of EFs at Time 2, $F(2, 166) = 27.26$, $p <.001$. Age, BAS and BPVS explained a total of 30.58% of the variance of the model. In step 3 of model 2a, symbolic maths at Time 1 was entered as final step and appeared as a significant predictor of EFs at Time 2 ($\beta = .537$), $F(1, 165) = 79.01$, $p < .001$, improving the model to explain 52.77% of the variance. We note here that this robust reciprocal relation between maths and EFs could only be tested because of our fully balanced design.

We then assessed whether Symbolic Maths at Time 1 predicted growth in EFs. Step 3 of model 2b consisted of the model as step 2 (age, BAS and BPVS as predictors) and instead of adding symbolic maths T1 as predictor (as in model 2a), the autoregressor EFs at T1 was entered. EFs at Time 1 as autoregressor improved the model fit significantly (*β = .641)*, $F(1, 165) = 136.1$, $p < .001$, with 61.72% of the variance explained by the model with age, BAS, BPVS and EFs at Time 1 as predictor of EFs at Time 2. As a final step to the autoregressive model 2b, symbolic maths at Time 1 was entered, which did not improve the model fit (*β = .090)*. This model explained 61.73% of the variance of the model, $F(1, 164) = 1.02$, $p = .315$.

Cross-lagged model

As a whole, the regression models reported above suggest that executive skills predict outcomes in symbolic maths skills 5 months later and their growth over that time period, but also that symbolic maths skills predict EF scores 5 months later, albeit not growth in EF from Time 1 to Time 2. The most complete model assessing the bidirectional relations between EFs and symbolic maths is a cross-lagged model that captures the inter-correlations between latent factors at Time 1 and Time 2, and tests for the extent to which they each predict each other longitudinally. The latent variables EFs and symbolic maths were entered in a cross-lagged regression model to examine whether EFs at Time 1 predicted symbolic maths at Time 2 as well as symbolic maths at Time 1 predicting EFs at Time 2 (model 3a), controlling for concurrent relations across measures. Age at Time 1 and the raw scores of BAS and BPVS were entered as control measures. However, fit indices indicate that this model was a poor fit for the data, Robust CFI = 0.720, Robust TLI = 0.370, Robust RMSEA = 0.389).

Furthermore, an autoregressive cross-lagged regression model was used to investigate whether EFs at Time 1 predicted growth in symbolic maths from Time 1 to Time 2 and if symbolic maths at Time 1 predicted growth in EF from Time 1 to Time 2 (model 3b), accounting for cross-factor correlations at each time-point. Age at Time 1 and raw scores of the BAS and BPVS were entered as control measures. Again, this model did not fit the data well, Robust CFI = 0.874, Robust TLI = 0.621, Robust RMSEA = 0.301.

As both cross-lagged regression models did not fit the data well, they will not be described further, but the models are displayed in the supplementary materials (Figures S2 and S3). Here, we provide more information on possible sources of poor model fit, supplemented by further statistics in the Supplementary Materials. We used Exploratory Factor Analyses (EFA) to get a better understanding of all the measures and their factor loadings, including the magnitude comparison measure, as this measure was not loading onto the domain-specific maths skill factor as we had hypothesised. The EFA models and the factor loadings for both Time 1 and Time 2 are presented in the Supplementary Materials (Table S3). Of note, the EFA at Time 1 showed that the Stroop measure loaded onto the maths factor instead of the EF factor (Figure S1). Furthermore, the magnitude comparison measure loaded onto both factors ($\beta > 0.3$; maths and EFs). However, in the EFA at Time 2, both the Stroop measure and the magnitude comparison measure only load onto the EF factor. We do not have a strong explanation for these unexpected findings and variable performance for the magnitude comparison task and Stroop task, although there already exists a literature that questions whether magnitude comparison is a pure measure of numerical skills, or whether it intrinsically calls on inhibitory control, as well as numerical skills, especially in very young children (Gilmore et al., 2013; Merkley et al., 2016). In addition, these changes in factor-loadings for the magnitude comparison task and the Animal Stroop task suggest that the structure of latent EF and maths factors at these two time-points in the preschool period, albeit based on previous literature in this area, is rather fragile and needs further investigation. We shall return to these in our discussion of findings outside our pre-registered hypotheses.

Discussion

The main aim of this study was to explore the interplay of domain-general and domain-specific cognitive skills in the development of early numeracy prior to formal schooling through a fully balanced longitudinal design (i.e., with an equal focus on foundational domain-general and domain-specific cognitive skills). Our first pre-registered hypothesis proposed that domain-general and domain-specific cognitive skills would load respectively onto two separable factors. This was indeed partly the case, when taking into account that the magnitude comparison task loaded onto a separate third factor. Our second pre-registered hypothesis stipulated that domain-general cognitive skills would predict maths skills longitudinally and predict growth in maths. This is indeed what we found, supporting abundant prior research (Blair & Razza, 2007; Bull et al., 2011; Bull & Scerif, 2001; Clark et al., 2013; Schmitt et al., 2017; Welsh et al., 2010; Ribner et al., 2020), although here we add to the existing literature because, first, we investigate change in children as young as 3 to 4 years of age, prior to the onset of formal education and, second, we study executive skills as a predictor of growth for foundational maths skills, unlike multi-componential classroom based achievement measures. These findings are also convergent with recently published findings in a large sample of American kindergarten students showing that EFs measured at the start of kindergarten (4-5 years of age) predicted growth in maths achievement over that same school year (Ribner, 2020).

For our third pre-registered hypothesis, we predicted that maths skills would predict individual differences in executive skills and growth in executive skills, consistent with mutualistic (Peng & Kievit, 2020) and transactional models (Miller-Cotto & Byrnes, 2019) of the interplay between domain-general cognitive skills and domain-specific educationally relevant outcomes. This hypothesis was only partially met. We found that number skills measured at Time 1 significantly correlated with EF composite at Time 2. However, when including EFs measured at Time 1 in the model as an autoregressor, we found that numeracy did not predict the change in EFs from Time 1 and Time 2. Of note, we did find stable relations between EFs and maths at both time points, and reciprocal predictions in terms of performance outcomes: children with poorer EFs at Time 1 had poorer maths at Time 2, a commonly reported finding, but, also, children with poorer maths at Time 1 had poorer EFs at Time 2, a rarely tested and / or reported finding. However, we found an asymmetry in these reciprocal relations, with EFs alone predicting change in maths over the five months of the study. The fact that maths symbolic skills did not predict growth in executive skills was not consistent with our third pre-registered hypothesis, nor is it consistent with previous findings of bidirectional associations between EF and overall maths achievement in this age group (e.g. Fuhs et al., 2014, although note the non-significant prediction from Applied Problems and Quantitative Concepts to EF growth in kindergarten children, who are more comparable to our younger target age group). It is therefore possible that bidirectional associations between domain-general cognitive abilities and foundational mathematics skill do exist but that we failed to capture them with the measures used in this study.

We now expand on both the predicted and unexpected findings. We had predicted that domain-general and domain-specific cognitive skills correlate strongly with each other concurrently and longitudinally even in preschool children. Here we highlighted that the relation between EFs and early numeracy holds for foundational symbolic skills (e.g., counting, cardinal number knowledge) and not only the academic achievement measures that were mostly investigated in prior research (Bull et al., 2011; Clark et al., 2013; Fuhs et al., 2014; Van der Ven et al., 2012; Welsh et al., 2010). The distinction between foundational and achievement measures is noteworthy, as EFs may contribute to achievement score as a limiting factor on test scores (e.g., by limiting children's ability to perform well on any challenging task), but they must instead play a greater learning-related role, because they predict foundational skills and their growth, especially prior to the onset of formal education and formal testing. Indeed, several ways in which domain-general skills could contribute to the development of domain-specific skills have previously been suggested. For instance Fuhs et al. (2014) suggested that EFs can assist children in their maths skills by helping them to switch between maths strategies when task demands change, maintain information in the working memory when solving numerical problems and inhibiting distracting information and unfavourable strategies. Moreover, domain-specific skills remain predicted by EFs over time as maths skills continue to build on one another, so that new activities such as calculations and word problems will continue to rely on EFs (Welsh et al., 2010). Furthermore, Fuhs, Hornburg, and McNeil (2016) later demonstrated that in slightly older children (6-year-olds) early number skills, in particular number set identification, fully mediated the link between EFs and maths achievement, while EFs did not mediate the link between early number skills and maths achievement. The authors suggested that one way in which EFs and maths achievement are related is through early foundational number skills.

The current study extends these findings by demonstrating that early foundational number skills and early domain-general skills such as EFs and attention are correlated both concurrently and longitudinally. A possible explanation for these robust correlations is that the demands of maths activities might enable children to cope with executive demands better, such as the demands on maintenance in working memory and on attentional focus. It may also be that some of the earliest-developing numerical abilities are intrinsically linked with attentional focus from the beginning: even in infancy, numerosity recognition and comparison tasks may require not only number awareness but selective attention to the number-relevant attributes of the stimulus (Merkley, Scerif, & Ansari, 2017 - in response to Leibovich et al., 2017).

Of note, however, foundational symbolic skills were correlated with domain-general skills longitudinally, but they did not predict their growth. This is at odds with mutualistic and transactional models proposed for older children (Kievit et al., 2019, from 6 years of age; Miller-Cotto & Byrnes, 2019, from 5 years of age). There is no clear explanation for this relative asymmetry in our data and for not fully meeting our third hypothesis. Here, we speculate with three non-mutually exclusive proposals. The first focuses on the age of our participants. Indeed, in a meta-analysis of 680 studies on the relations between nonverbal reasoning and reading/mathematics across a wide range of age-groups (3–80 years old; Peng, Wang, Wang, & Lin, 2019), bidirectional relations across domains become increasingly stronger over age, with weakest relations in the youngest children. This logic has not been applied to the relations between maths and EFs. A second possibility is more focused on the cognitive demands of EFs and maths skills: over the course of learning and development the demands and sophistication of maths activities change and may begin to contribute to improvements in EF skills: it is possible that relations between maths and EFs become more symmetrically bidirectional, as EFs load specific to maths tasks increase when children do demanding, multi-digit calculation problems, for example. More specifically, it is possible that, at our target age, children grow most in their fundamental maths skills when they can pay attention/have good EFs, but that when maths strategies become more sophisticated and demanding on fluency, they provide further mechanisms through which EFs can grow. Further studies will need to investigate this question by testing bidirectional growth in older children to those studied here. In the interim, it is worth noting that an equally asymmetrical pattern was also found in kindergarten children in Fuhs et al. (2014), with specific maths achievement measures (i.e., Woodcock-Johnson III – Test of Achievement, Applied Problems and Quantitative Concepts) predicting EFs, but not growth in EFs. Welsh et al. (2010), on the other hand, found that early maths achievement (i.e., Woodcock-Johnson III – Test of Achievement) also predicted growth in EFs in preschool children. Reasons for this discrepancy (e.g., target age group, different performance demands; focus on subscales as opposed to overall scores; sample characteristics, as Welsh et al., focused on a sample from low-income communities) remain unclear and need to be investigated further and longitudinally. Moreover, Finch (2019) showed that kindergarten education more broadly was associated with growth in EFs. Although there are some inconsistencies across these studies, as a whole they call for further exploration of bidirectional relations, in fully balanced longitudinal designs that include foundational skills (as in our case), in addition to achievement measures (Fuhs et al., 2014; Welsh et al., 2010). A third possible explanation for not fully meeting our bidirectional hypothesis is dependent on the poor statistical fit of our full cross-lagged model. As we document through our exploratory analyses, at this age and with our measures, maths and EF measures cluster into two separable components, but individual measures (magnitude comparison, animal Stroop) are not stably associated with the same latent factor at both time points. We therefore return to the factor structure of executive skills and maths skills, our first pre-registered hypothesis.

We had hypothesised that EF and maths skills would load onto two separable factors at this age. This was the case, and indeed an alternative unitary CFA model encompassing all early numerical and cognitive skills showed that the data was best fit by a model separating EFs and numerical skills (see supplementary material for more details). However, our findings were further nuanced. The non-symbolic and symbolic maths measures best loaded onto separate factors, which has previously been debated in the literature (Friso-van den Bos et al., 2014; Kolkman et al., 2013; Xenidou-Dervou et al., 2013). Furthermore, Exploratory Factor Analyses (please see supplementary material) suggested that the magnitude comparison task, at least as used in this study, reflects EFs as well as early numerical skills, in line with current ongoing debates (e.g., Gebuis, Cohen Kadosh, & Gevers, 2016; Wilkey, Pollack, & Price, 2018; Keller & Libertus, 2015). Indeed, the magnitude comparison measure loaded onto both factors at Time 1, and only on the EF factor at Time 2, contributing to poor model fit for the full cross-lagged model. This fits in with the current debate about whether magnitude comparisons purely measure numerical judgements or whether inhibitory control might be reflected in the task (e.g., Merkley et al., 2016). This may be because, in the magnitude comparison task, just like in the Stroop task, children are required to inhibit visual information in incongruent trials, thus demonstrating the resemblance of the magnitude comparison task to EF tasks. However, we are certainly not well placed to resolve this debate here: an important limitation of the current study was the inclusion of only one non-symbolic maths measure. By including multiple number sense measures, future research might allow for the investigation of the directionalities between domain-general cognitive skills such as EFs and attention, early symbolic maths skills and early non-symbolic maths skills.

Moreover, we thank two anonymous reviewers for highlighting how, beyond questioning the executive and numerical demands of the magnitude comparison and the Animal Stroop task, the exploratory factor analyses findings suggest fragility / possible change in the latent maths and EF constructs measured at this young age. Here, we maintained our focus on the proposed hypothesis and registered analysis of unidirectional vs. bidirectional interplay between EFs and maths longitudinally, with a lesser focus on explaining why the measures do not clearly adhere to this model. However, future work needs to explore the question of how magnitude comparison loads with other measures across time and factors further, as well as the fragility of the underlying latent factors in these young children, for which formal tests of measurement invariance will be extremely helpful. We think that this change in cross-loadings, even over the course of just 5 months, reflects real life individual differences and is maybe a function of early developmental change or instability in these constructs.

A further limitation of this study is that the sample is fairly homogenous with respect to race and economic advantage. Future research should investigate development of executive functions in more diverse contexts, as culture and education systems influence cognitive development from executive functions to academic achievement (e.g., Finch, 2019; Howard et al., 2020; Lewis et al., 2009; Morrison et al., 2020; Welsh et al., 2010; Wolf & McCoy, 2019).

Are there educational implications of these data? Education has a strong influence on cognitive development (Morrison et al., 2019) and cognitive skills and academic achievement reciprocally grow over development (Peng & Kievit, 2019). Our findings on bidirectionality here are tentative. Future research should investigate causal mechanisms more explicitly within the context of well controlled educational interventions: if educational strategies that brought together EF and maths skills training, were more successful in improving both EFs and maths than separate interventions, bidirectional models would gain further traction. An educational approach combining domain-general and domain-specific skills could also prove more beneficial than specific attempts to improve domain-general skills, such as most EF training studies at present. Although training EFs in preschool through cognitive training paradigms seems possible (e.g., Blakey & Carroll, 2015; Thorell, Lindqvist, Bergman Nutley, Bohlin, & Klingberg, 2009), support for long-lasting far transfer to maths improvement is lacking (Melby-Lervåg & Hulme, 2013) The lack of transfer could depend on the fact that number-specific EFs are more strongly related to maths ability than EFs measured by tasks that do not include numerically-relevant stimuli (Fuhs, Nesbitt, & O'Rear, 2018; Gilmore, Keeble, Richardson, & Cragg, 2015; Wilkey et al., 2018). For this reason, ways to bring domain-general and domain-specific skills together are starting to emerge (e.g., Clements & Sarama, 2014; Howard, Powell, Vasseleu, Johnstone, & Melhuish, 2017; Howard, Vasseleu, Neilsen-Hewett, & Cliff, 2018; Ramani et al., 2017) and could be evaluated in preschool environments to investigate whether these promote the development of early numeracy skills. Early years educators have expressed desire for more training in early maths instruction (von

Spreckelsen et al., 2019; Youmans et al., 2018) and it remains to be tested whether teaching teachers to support both number-specific and domain-general cognitive skills in the early years could lead to improvements in early years maths outcomes in the UK. Preschool intervention studies targeting EFs and mathematics in North American children have provided some preliminary but conflicting evidence in this regard (Clements et al., 2020; cf. McClelland et al., 2019). Specifically, McClelland and colleagues (2019) found that children in a self-regulation intervention and children in a self-regulation and math combined intervention showed improvements in math compared to a business-as-usual control. In contrast, Clements et al. (2020) did not find evidence for differentially greater effectiveness of a combined intervention. More research is needed to understand the relationships between EFs and mathematics and investigate the implementation of classroom interventions.

In sum, our findings show that foundational domain-general and domain-specific cognitive skills in preschool to later maths achievement correlate strongly, although they are asymmetric with regards to growth in EFs. Our results indicate that domain-general and domain-specific skills in preschool children are highly and stably correlated at multiple time points and that some measures (such as magnitude comparisons and resolving Stroop interference) relate to both domain-general and domain-specific factors, making it difficult to disentangle domain-specific from domain-general skills in children as young as preschoolers.

References

Amso, D., & Scerif, G. (2015). The attentive brain: insights from developmental cognitive neuroscience. *Nature Reviews Neuroscience*, *16*(10), 606-619.

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Blakey, E., & Carroll, D. J. (2015). A Short Executive Function Training Program Improves Preschoolers' Working Memory. *Frontiers in Psychology*, *6*, 1827. https://doi.org/10.3389/fpsyg.2015.01827

Bull, R., Espy, K. A., Wiebe, S. A., Sheffield, T. D., & Nelson, J. M. (2011). Using confirmatory factor analysis to understand executive control in preschool children: sources of variation in emergent mathematic achievement. *Developmental Science*, *14*(4), 679–692. https://doi.org/10.1111/j.1467-7687.2010.01012.x

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: inhibition, switching, and working memory. *Developmental Neuropsychology*, *19*(3), 273–293. https://doi.org/10.1207/S15326942DN1903_3

Chu, F. W., vanMarle, K., & Geary, D. C. (2015). Early numerical foundations of young children's mathematical development. *Journal of Experimental Child Psychology*, *132*, 205–212. https://doi.org/10.1016/j.jecp.2015.01.006

Clark, C. A. C., Sheffield, T. D., Wiebe, S. A., & Espy, K. A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child Development*, *84*(2), 662–677. https://doi.org/10.1111/j.1467-8624.2012.01854.x

Clements, D. H., & Sarama, J. (2014). *Learning and teaching early math : the learning trajectories approach* (second edi). Routledge.

Clements, D. H., Sarama, J., & Germeroth, C. (2016). Learning executive function and early mathematics: Directions of causal relations. *Early Childhood Research Quarterly*, *36*, 79–90. https://doi.org/10.1016/j.ecresq.2015.12.009

Clements, D. H., Sarama, J., Layzer, C., Unlu, F., & Fesler, L. (2020). Effects on mathematics and executive function of a mathematics and play intervention versus mathematics alone. *Journal for Research in Mathematics Education*, *51*(3), 301-333.

Dalmaijer, E. S., Van der Stigchel, S., Nijboer, T., Cornelissen, T., & Husain, M. (2015). CancellationTools: All-in-one software for administration and analysis of cancellation tasks. Behavior research methods, 47(4), 1065–1075. https://doi.org/10.3758/s13428-014-0522-7

Doherty, B. R., Charman, T., Johnson, M. H., Scerif, G., Gliga, T., & BASIS Team (2018). Visual search and autism symptoms: What young children search for and co-occurring ADHD matter. Developmental science, 21(5), e12661. https://doi.org/10.1111/desc.12661

Dowker, A. (2005). Early Identification and Intervention for Students With Mathematics Difficulties. *Journal of Learning Disabilities*, *38*(4), 324–332. https://doi.org/10.1177/00222194050380040801

Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Developmental Science*, *11*(5), 650–654. https://doi.org/10.1111/j.1467-7687.2008.00713.x

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

English Indices of Deprivation. (2015). *English Indices of Deprivation 2015: Postcode*

Lookup. http://imd-by-postcode.opendatacommunities.org/

Finch, J. E. (2019). Do Schools Promote Executive Functions? Differential Working Memory Growth Across School-Year and Summer Months. *AERA Open*, *5*(2), 2332858419848443.

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

Fuhs, M. W., Hornburg, C. B., & McNeil, N. M. (2016). Specific early number skills mediate the association between executive functioning skills and mathematics achievement. *Developmental Psychology*, *52*(8), 1217–1235. https://doi.org/10.1037/dev0000145

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. https://doi.org/10.1111/desc.12013

Fuhs, M. W., Nesbitt, K. T., Farran, D. C., & Dong, N. (2014). Longitudinal associations between executive functioning and academic skills across content areas. *Developmental Psychology*, *50*(6), 1698–1709. https://doi.org/10.1037/a0036633

Fuhs, M. W., Nesbitt, K. T., & O'Rear, C. D. (2018). Approximate number system task performance: Associations with domain-general and domain-specific cognitive skills in young children. *Journal of Numerical Cognition*, *4*(3), 590–612. https://doi.org/10.5964/jnc.v4i3.141

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychological bulletin*, 134(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*(6), 1539–1552.

https://doi.org/10.1037/a0025510

Gebuis, T., Cohen Kadosh, R., & Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. *Acta Psychologica*, *171*, 17–35. https://doi.org/10.1016/j.actpsy.2016.09.003

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., Simms, V., & Inglis, M. (2013). Individual Differences in Inhibitory Control, Not Non-Verbal Number Acuity, Correlate with Mathematics Achievement. *Plos One*, *8*(6). https://doi.org/10.1371/journal.pone.0067374

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

Gilmore, C., Keeble, S., Richardson, S., & Cragg, L. (2015). The role of cognitive inhibition in different components of arithmetic. *Zdm*, *2010*. https://doi.org/10.1007/s11858-014-0659-y

Gimbert, F., Camos, V., Gentaz, E., & Mazens, K. (2019). What predicts mathematics achievement? Developmental change in 5- and 7-year-old children. *Journal of Experimental Child Psychology*, *178*, 104–120. https://doi.org/10.1016/J.JECP.2018.09.013

Hendry, A., Jones, EJH., & Charman, T. (2016). Executive function in the first three years of life: Precursors, predictors and patterns. *Developmental Review*, 42, 1-33, https://doi.org/10.1016/j.dr.2016.06.005.

Hassinger-Das, B., Jordan, N. C., Glutting, J., Irwin, C., & Dyson, N. (2014). Domain-general mediators of the relation between kindergarten number sense and first-grade mathematics achievement. *Journal of Experimental Child Psychology*, *118*, 78–92. https://doi.org/10.1016/j.jecp.2013.09.008

Howard, S. J., & Melhuish, E. (2017). An Early Years Toolbox for Assessing Early Executive Function, Language, Self-Regulation, and Social Development: Validity, Reliability, and Preliminary Norms. *Journal of Psychoeducational Assessment*, *35*(3), 255–275. https://doi.org/10.1177/0734282916633009

Howard, S. J., Powell, T., Vasseleu, E., Johnstone, S., & Melhuish, E. (2017). Enhancing Preschoolers' Executive Functions Through Embedding Cognitive Activities in Shared Book Reading. *Educational Psychology Review*, *29*(1), 153–174. https://doi.org/10.1007/s10648-016-9364-4

Howard, S. J., Vasseleu, E., Neilsen-Hewett, C., & Cliff, K. (2018). Evaluation of the Preschool Situational Self-Regulation Toolkit (PRSIST) Program for Supporting children's early self-regulation development: study protocol for a cluster randomized controlled trial. *Trials*, *19*(1), 64. https://doi.org/10.1186/s13063-018-2455-4

Howard, S. J., Cook, C. J., Everts, L., Melhuish, E., Scerif, G., Norris, S., Twine, R., Kahn, K., & Draper, C. E. (2020). Challenging socioeconomic status: A cross-cultural comparison of early executive function. *Developmental science, 23(1*), e12854. https://doi.org/10.1111/desc.12854

Keller, L., & Libertus, M. (2015). Inhibitory control may not explain the link between approximation and math abilities in kindergarteners from middle class families. *Frontiers in Psychology*, *6*, 685. https://doi.org/10.3389/fpsyg.2015.00685

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

Köse, A. (2014). The effect of misssing data handling methods on goodness of fit indices in confirmatory factor analysis. *Educational Research and Reviews*, *9*(8), 208–215. https://doi.org/10.5897/ERR2014.1709

Lau, N. T., Merkley, R., Tremblay, P. F., Zheng, S., De Jesus, S., & Ansari, D. (in press). Kindergarteners' Symbolic Number Abilities Predict Non-Symbolic Number Abilities and Math Achievement in Grade 1. *Developmental Psychology*. https://doi.org/10.31234/osf.io/xpe9c

Lee, K., Bull, R., & Ho, R. M. H. (2013). Developmental changes in executive functioning. *Child Development*, *84*(6), 1933–1953. https://doi.org/10.1111/cdev.12096

Lehto, J. E., Juujarvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. The British Journal of Developmental Psychology, 21, 59. https://doi.org/10.1348/026151003321164627

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS ONE*, *6*(9), e23749. https://doi.org/10.1371/journal.pone.0023749

McKinnon, R. D., & Blair, C. (2019). Bidirectional relations among executive function, teacher–child relationships, and early reading and math achievement: A cross-lagged panel analysis. *Early Childhood Research Quarterly*, *46*, 152–165. https://doi.org/10.1016/j.ecresq.2018.03.011

McClelland, M. M., Tominey, S. L., Schmitt, S. A., Hatfield, B., Purpura, D., Gonzales, C., & Tracy, A. (2019). Red Light, Purple Light! Results of an intervention to promote school readiness for children from low-income backgrounds. *Frontiers in Psychology*, *10*, 2365.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291. https://doi.org/10.1037/a0028228

Merkley, R., & Ansari, D. (2016). Why numerical symbols count in the development of mathematical skills: evidence from brain and behavior. *Current Opinion in Behavioral*

Sciences, *10*, 14–20. https://doi.org/10.1016/J.COBEHA.2016.04.006

Merkley, R., Thompson, J., & Scerif, G. (2016). Of Huge Mice and Tiny Elephants: Exploring the Relationship Between Inhibitory Processes and Preschool Math Skills. *Frontiers in Psychology*, *6*, 1903. https://doi.org/10.3389/fpsyg.2015.01903

Merkley, R., Matusz, P. J., & Scerif, G. (2018). The control of selective attention and emerging mathematical cognition: Beyond unidirectional influences. In *Heterogeneity of function in numerical cognition* (pp. 111-126).

Merkley, R., Scerif, G., & Ansari, D. (2017). What is the precise role of cognitive control in the development of a sense of number?. *Behavioral and Brain Sciences*, *40*.

Miller-Cotto, D., & Byrnes, J. P. (2019). What's the best way to characterize the relationship between working memory and achievement?: An initial examination of competing theories. *Journal of Educational Psychology*.

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current directions in psychological science,* 21(1), 8–14. https://doi.org/10.1177/0963721411429458

Morrison, F. J., Kim, M. H., Connor, C. M., & Grammer, J. K. (2019). The Causal Impact of Schooling on Children's Development: Lessons for Developmental Science. *Current Directions in Psychological Science*, *28*(5), 441-449.

Mulder, H., Verhagen, J., Van der Ven, S. H. G., Slot, P. L., & Leseman, P. P. M. (2017). Early Executive Function at Age Two Predicts Emergent Mathematics and Literacy at Age Five. *Frontiers in Psychology*, *8*, 1706. https://doi.org/10.3389/fpsyg.2017.01706

Nosworthy, N. (2013). An Investigation of the Association Between Arithmetic Achievement and Symbolic and Nonsymbolic Magnitude Processing in 5-9 Year-old Children: Evidence from a Paper-and-pencil Test [the University of Western Ontario]. In *doctoral dissertation*. https://ir.lib.uwo.ca/etd/1387

Park, J., & Brannon, E. M. (2014). Improving arithmetic performance with number sense training: an investigation of underlying mechanism. *Cognition*, *133*(1), 188–200. https://doi.org/10.1016/j.cognition.2014.06.011

Peng, P., & Kievit, R. A. (2020). The development of academic achievement and cognitive abilities: A bidirectional perspective. *Child Development Perspectives*, *14*(1), 15-20.

Peng, P., Wang, T., Wang, C., & Lin, X. (2019). A meta-analysis on the relation between fluid intelligence and reading/mathematics: Effects of tasks, age, and social economics status. *Psychological Bulletin*, *145*(2), 189.

Prager, E. O., Sera, M. D., & Carlson, S. M. (2016). Executive function and magnitude skills in preschool children. *Journal of Experimental Child Psychology*, *147*, 126–139. https://doi.org/10.1016/j.jecp.2016.01.002

Price, G. R., & Wilkey, E. D. (2018). Cognitive mechanisms underlying the relation between nonsymbolic and symbolic magnitude processing and their relation to math. *Cognitive Development*, *44*(Supplement C), 139–149. https://doi.org/10.1016/j.cogdev.2017.09.003

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. https://doi.org/10.1037/a0031753

Purpura, D. J., Schmitt, S. A., & Ganley, C. M. (2017). Foundations of mathematics and literacy: The role of executive functioning components. *Journal of Experimental Child Psychology*, *153*, 15-34.

Ramani, G. B., Jaeggi, S. M., Daubert, E. N., & Buschkuehl, M. (2017). Domain-specific and domain-general training to improve kindergarten children's mathematics. *Journal of Numerical Cognition*, *3*(2), 468-495.

Ribner, A. D. (2020). Executive function facilitates learning from math instruction in kindergarten: Evidence from the ECLS-K. *Learning and Instruction*, *65*, 101251.

Steele, A., Karmiloff-Smith, A., Cornish, K., & Scerif, G. (2012). The multiple subfunctions of attention: differential developmental gateways to literacy and numeracy. *Child Development*, *83*(6), 2028–2041. https://doi.org/10.1111/j.1467-8624.2012.01809.x

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651

Thorell, L. B., Lindqvist, S., Bergman Nutley, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12*(1), 106–113. https://doi.org/10.1111/j.1467-7687.2008.00745.x

Van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., & Leseman, P. P. M. (2012). The development of executive functions and early mathematics: a dynamic relationship. *The British Journal of Educational Psychology*, *82*(Pt 1), 100–119. https://doi.org/10.1111/j.2044-8279.2011.02035.x

von Spreckelsen, M., Dove, E., Coolen, I., Mills, A., Dowker, A., Sylva, K., ... & Scerif, G. (2019). Let's Talk About Maths: The Role of Observed “Maths-Talk” and Maths Provisions in Preschoolers' Numeracy. *Mind, Brain, and Education*, *13*(4), 326-340.

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., & Nelson, K. E. (2010). The Development of Cognitive Skills and Gains in Academic School Readiness for Children from Low-Income Families. *Journal of Educational Psychology, 102*(1), 43–53. https://doi.org/10.1037/a0016738

Wilkey, E. D., Pollack, C., & Price, G. R. (2018). Dyscalculia and Typical Math Achievement Are Associated With Individual Differences in Number-Specific Executive Function. *Child Development*, *00*(0), 1–24.

https://doi.org/10.1111/cdev.13194

Wolf, S., & McCoy, D. C. (2019). The role of executive function and social-emotional skills in the development of literacy and numeracy during preschool: a cross-lagged longitudinal study. *Developmental science, 22(*4), e12800. https://doi.org/10.1111/desc.12800

Woods, A. J., Göksun, T., Chatterjee, A., Zelonis, S., Mehta, A., & Smith, S. E. (2013). The development of organized visual search. Acta psychologica, 143(2), 191–199. https://doi.org/10.1016/j.actpsy.2013.03.008

Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cognitive Psychology*, *24*(2), 220–251. https://doi.org/10.1016/0010-0285(92)90008-P

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

Youmans, A., Coombs, A., & Colgan, L. (2018). Early Childhood Educators' and Teachers' Early Mathematics Education Knowledge, Beliefs, and Pedagogy. *Canadian Journal of Education*, *41*(4), 1079-1104.