

Empirical Research



Longitudinal Predictors of Conceptual Understanding of Arithmetic Principles

Silke M. Göbel^{1,2} , Karin Landerl³ , Arne O. Lervåg^{2,4} 

[1] Department of Psychology, University of York, York, United Kingdom. [2] Centre for Research on Equality in Education (CREATE), University of Oslo, Oslo, Norway. [3] Department of Psychology, University of Graz, Graz, Austria. [4] Department of Education, University of Oslo, Oslo, Norway.

Journal of Numerical Cognition, 2026, Vol. 12, Article e19523, <https://doi.org/10.5964/jnc.19523>

Received: 2025-08-27 • Accepted: 2025-12-04 • Published (VoR): 2026-03-19

Handling Editors: Daniel B. Berch, School of Education and Human Development, University of Virginia, Charlottesville, VA, USA; Bert De Smedt, Faculty of Psychology and Educational Sciences, University of Leuven, Leuven, Belgium

Corresponding Author: Silke M. Göbel, Heslington, York YO10 4AD, United Kingdom. Tel: +44 1904 322872. E-mail: silke.goebel@york.ac.uk

Related: This article is part of the JNC Special Thematic Collection “Advancing Our Knowledge of Mathematical Understanding”, Guest Editors: Daniel B. Berch, Jo-Anne LeFevre, Bert De Smedt, & Helena P. Osana. <https://doi.org/10.5964/jnc.arco1>

Supplementary Materials: Data [see [Index of Supplementary Materials](#)]



Abstract

A bidirectional relationship between conceptual and procedural understanding in the development of arithmetic skills has often been reported. We investigated whether domain-specific longitudinal predictors of procedural arithmetic performance at the beginning of primary school also predict conceptual understanding two years later. We assessed conceptual and procedural understanding of arithmetic and mathematical reasoning in 195 UK children (mean age 8 years 2 months) in Year 3. Conceptual understanding was defined as children’s understanding of principles underlying arithmetic procedures. Performance on a speeded arithmetic task was taken as an indicator of children’s procedural understanding of arithmetic. The same children had been assessed in Year 1 on potential cognitive and numerical predictors including number transcoding, symbolic and non-symbolic magnitude comparison, arithmetic performance, verbal and visuo-spatial working memory, and non-verbal reasoning. A structural equation model including arithmetic performance, number transcoding and non-verbal cognitive skills measured in Year 1 predicted 33% of the variance in conceptual understanding in Year 3. Arithmetic performance and number transcoding



This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

in Year 1 were also significant longitudinal predictors of both procedural arithmetic understanding and mathematical reasoning in Year 3. When we ran a second structural equation model without arithmetic performance in Year 1, number transcoding and non-verbal cognitive skills remained the only significant longitudinal predictors of conceptual understanding in Year 3. Our study highlights substantial similarities as well as some differences in the longitudinal predictors of conceptual versus procedural understanding of arithmetic in early primary school.

Keywords

commutativity, inversion, subtraction complement, explanations, transcoding, mathematical reasoning

The development of mathematical understanding is often seen as an important goal of mathematical learning. In the current study we measured children's mathematical understanding along with their arithmetic fact retrieval in Year 3 and investigated which numerical and cognitive skills in Year 1 predict individual differences in mathematical understanding and performance two years later.

Conceptual Understanding

Definition and Measurement

Conceptual understanding has been described as the *understanding of why* and often in contrast to procedural understanding (which is the *understanding of how* to do something; [Hiebert & Lefevre, 1986](#)). Procedural understanding is typically defined as 'understanding of procedures for solving problems, such as the step-by-step algorithms that children are taught in school' ([Braithwaite & Sprague, 2021](#)). Conceptual understanding is the understanding of the principles and relationships within a particular area. Conceptual understanding can help children for example to decide which procedure is the best to use in a particular situation. It can be used to sense-check a given solution and might be useful when encountering novel problem types ([Gilmore, 2023](#)).

That there is no consistent definition of conceptual mathematical understanding might be because conceptual understanding in mathematics captures various different aspects. In their review of conceptual understanding in mathematics [Crooks and Alibali \(2014\)](#) list six different types of definitions for conceptual understanding they found in the literature. Those definition types are 1. understanding of relationships within a domain, 2. understanding of general rules, facts and definitions (general principle understanding), 3. understanding of principles underlying procedures, 4. category understanding, 5. symbol understanding and 6. domain-structure understanding. Research on conceptual understanding in mathematics has been carried out on diverse, often quite separate, mathematical topics: for example, cardinality understanding ([Sarnecka & Wright, 2013](#)), understanding of arithmetic principles (including inversion; [Gilmore & Papadatou-Pastou, 2009](#)), the understanding of equivalence ([McNeil et al., 2019](#)) and

fraction understanding (Siegler & Lortie-Forgues, 2015). One key question is how to measure conceptual understanding based on the child's behaviour. Bisanz and LeFevre (1992) suggested three different ways of measuring conceptual understanding: first by measuring the use of the corresponding procedures to solve problems, second by asking children for an explicit justification of procedures mentioning the principles (which we will call explanations), and third by requiring an evaluation from children whether these procedures can be applied or have been applied correctly (which we will call judgements). The first suggestion actually measures the child's procedural understanding. These three abilities do not always go hand in hand. A child might be able to justify a procedure without being able to apply it (Gelman & Gallistel, 1978), or they might apply the procedure but can not justify why they used it (Sophian, 1995). Asking the child to give an explanation that justifies a particular procedure gives clear evidence of explicit understanding of the underlying concept. An appropriate judgement does not require the use of the procedure and also does not need explicit verbalisation of a principle. Because the last method has the lowest cognitive demands on a child we decided to use judgements as our main measure of conceptual understanding. Bisanz and LeFevre (1992) advise to use at least two of the measures together, thus on a subset of the items we also asked children to give an explanation for their judgement.

Conceptual Understanding of Arithmetic Principles

The current paper focuses on the conceptual understanding of arithmetic principles related to addition and subtraction. Thus, we will limit our discussion of the definition and tasks used for conceptual understanding here to the conceptual understanding of arithmetic principles. In our study we will focus on three different types of arithmetic principles related to addition and subtraction that children acquire. *Commutativity* means that the order of the two operands in an addition problem does not matter, i.e. $a + b = c$ AND $b + a = c$. For subtraction problems, the *subtraction complement* gives the same answer, i.e., $c - a = b$ AND $c - b = a$. The *inversion principle* is about the inverse relationship between addition and subtraction, i.e. $a + b = c$ AND $c - a = b$. Several studies provide evidence that out of those three principles children find the commutativity principle comparatively easy (e.g., Canobi, 2004) and acquire this principle first. With increasing grade level many children then move from understanding only additive commutativity to also understanding principles related to subtraction, such as the subtraction complement principle (Canobi, 2004) and inversion (Canobi, 2005). For example, less than 15% of the 6-8 years olds in Canobi's 2004 study understood the inversion and subtraction complement principle, while 88% grasped the commutativity principle. In another study by Canobi (2005) 75% of the 7-year-olds were still struggling with the understanding of concepts involving subtraction, while for 8- and 9- year-old children in their study that figure dropped to about 50%. Understanding the inverse relationship between addition and subtraction (the inversion principle) has received particular interest in the literature

(e.g. Robinson & Dubé, 2009b). A meta-analysis by Gilmore and Papadatou-Pastou (2009) including 14 studies of the understanding of inversion with children aged 5 to 13 years showed reliable patterns of individual differences in children's understanding of the inversion principle. Surprisingly, in this meta-analysis neither children's age nor their grade level significantly affected the size of the inversion effect. However, this could be because different inversion task versions were used in different studies and for children in different school years. For example, while 4- and 5-year-old children already understood the inversion principle (e.g. Rasmussen et al., 2003) when presented with so-called three-term symbolic inverse problems (e.g., $5 + 2 - 2$), Canobi (2005) found that only 10% of their 5 year-olds understood inversion with two-term complementary items (e.g. $3 + 4 = 7$, $7 - 4 =$) while this was the case for 30% of the 6- and 7-year-olds in their study. Similarly, for three-term inverse problems 5 year-old children showed inversion understanding for large approximate arithmetic problems but not for exact arithmetic problems (Gilmore & Spelke, 2008).

Giving verbal explanations of concepts is in general more difficult for children than to judge whether they could use specific arithmetic concepts to solve arithmetic problems. For example, Canobi et al. (2003) tested 5- to 8-year-old children on their conceptual and procedural understanding of addition. In their study children, independent of their age, found it more difficult to provide accurate explanations for their judgements than to judge whether a particular concept (e.g., commutativity) could be used to solve a particular addition problem.

Longitudinal Predictors of Conceptual Understanding

With the exception of Andersson (2010) whose focus was on children with learning difficulties, there are hardly any studies that have compared longitudinally and within the same design the predictive power of a set of cognitive skills in early primary school for later conceptual understanding of arithmetic principles. Given the large inter-individual differences in children's conceptual understanding of arithmetic principles and the proposed foundational nature of conceptual understanding for mathematical learning, identifying early foundations seems an important endeavor. Both domain-specific numerical skills and domain-general cognitive skills might play important roles in the development of conceptual understanding of arithmetic principles.

Domain-Specific Numerical Skills

Magnitude Understanding — A key numerical skill for understanding addition and subtraction and the differences between them is magnitude understanding. Children have to grasp the magnitude of the operands and most children learn early that addition leads to larger and subtraction to smaller magnitudes (McCrink & Wynn, 2004). It has been proposed that children are born with an approximate number system (Dehaene, 2001; de Hevia et al., 2017), i.e. an ability to extract imprecise numerosity information

from collections of objects or sounds, that gets more accurate with age and experience (Halberda & Feigenson, 2008; Piazza et al., 2013). There is some evidence that pre-school non-symbolic number comparison predicts later performance in symbolic arithmetic (Libertus et al., 2011; Malone et al., 2021).

Once children start formal schooling number symbols become increasingly important and symbolic magnitude understanding becomes a stronger predictor of their arithmetic development (Schneider et al., 2017). In our study we are using symbolic and non-symbolic number comparison as a measure of children's magnitude understanding.

Multi-Digit Understanding – When children are acquiring number symbol knowledge they typically first learn single digits, but quickly move onto multi-digit strings. To grasp the meaning of multi-digit strings, an understanding of place-value, i.e. the fact that the value of a particular digit in a string is determined by its location in the string, is necessary. Arithmetic understanding entails an understanding of place value (Bisanz & LeFevre, 1992), particularly for arithmetic problems that cross decade boundaries and involve carrying or borrowing (Dresen et al., 2020). In addition, the conceptual understanding task we have used in this study uses multi-digit operands, thus we expect place-value understanding to be a predictor of conceptual understanding. In our study we measured place-value understanding with multi-digit number transcoding which has been shown to be a longitudinal predictor of arithmetic development (Banfi et al., 2022; Göbel et al., 2014; Malone et al., 2021).

We measured multi-digit number transcoding with three tasks, 1. number reading (reading aloud Arabic digit strings), 2. number writing (writing spoken number words down as Arabic digit strings) and 3. number identification (identifying the correct digit string from several choices after hearing a spoken number word), including multi-digit items outside the range of numbers focused on in the Year 1 curriculum. The most common errors in number reading and number writing in English-speaking children in Year 1 (see Steiner et al., 2021) are syntactic errors (additive composition, e.g., writing 600403 for 'six hundred forty-three'). Those errors, such as writing a 6-digit number for a spoken number word in the hundreds, demonstrate an incomplete place-value understanding. Thus, while multi-digit number transcoding is an indicator of children's ability to move between spoken number words and written Arabic digits, in our view it also reflects a rudimentary form of place-value understanding.

Arithmetic Skills – The relationship between conceptual understanding of arithmetic principles and children's performance on arithmetic tasks, often described as their procedural skills, has been researched extensively. For a long time the view that procedural skill development followed conceptual understanding was widely held (Hiebert & Lefevre, 1986; Rittle-Johnson et al., 2015). Several studies, for example, did not find a relationship between children's performance on an inversion task and their arithmetic

skills with quantities of a similar magnitude (Rasmussen et al., 2003). In a meta-analysis across 14 studies with children aged 5 to 13 years, Gilmore and Papadatou-Pastou (2009) found evidence for three different groups: children with good conceptual understanding and good arithmetic skills, children who performed poorly on both tasks and a third group of children who showed good understanding of the inversion principle but poor calculation skills. This last group is of particular interest because it highlights that it might be possible to develop good conceptual understanding (of inversion) before having good arithmetic skills.

In contrast, others have proposed that an understanding of arithmetic principles such as inversion only emerges after children have achieved a good level of addition and subtraction problem solving. Baroody (1999) for example reported that children showed a better inversion understanding for well-learned problems. In a cross-sectional study of year 1 to year 3 children by Canobi (2005) differences in conceptual understanding were related to the use of fact-retrieval based strategies for solving addition and subtraction problems.

More recent studies provide evidence for a bidirectional relation of procedural and conceptual skills (e.g., Rittle-Johnson et al., 2015). When 7- to 8-year-old children received structured procedural training their ability to provide conceptual explanations improved and the size of improvement in conceptual understanding was predicted by initial arithmetic performance (Canobi, 2009). Based on these studies we expected arithmetic performance in Year 1 to be a significant predictor of conceptual understanding of arithmetic principles in Year 3. Unfortunately, we had no measure of conceptual understanding in Year 1, and thus are unable to test a possible bidirectional relationship between procedural and conceptual arithmetic skills.

Domain-General Cognitive Skills

Non-Verbal Reasoning – Non-verbal reasoning is often measured as the ability to recognise and complete visuo-spatial patterns using reasoning. Children with higher general non-verbal cognitive skills might extract rules and regularities more easily, which is essential for developing conceptual understanding. In line with this argument, Andersson (2010) found a composite measure of verbal and non-verbal IQ was a significant predictor of conceptual understanding of the commutative and the inversion principles in 9 to 13 year-old children. Thus, we included a non-verbal measure of general cognitive skills to control for individual differences in non-verbal reasoning.

Verbal Working Memory (WM) – In order to access conceptual understanding, verbal WM might be crucial when retrieving relevant information stored in long-term memory (Cragg et al., 2017). However, verbal WM does not seem to be related to conceptual understanding of fractions (Jordan et al., 2013) and verbal WM measured as listening span also was not a significant longitudinal predictor of children's conceptual understanding

of the commutative and the inversion principles in Andersson (2010). In contrast, Cragg et al. (2017) found a significant concurrent relationship between verbal WM and conceptual understanding of arithmetic principles. Given that we are using Cragg et al. (2017)'s task in our study, we expected verbal WM to emerge as a significant predictor of conceptual understanding.

Visuo-Spatial Working Memory (WM) — Early differences in visuo-spatial WM have often been found to predict children's later procedural arithmetic skills (Peng et al., 2016). This could be due to employing a mental number line for performing addition and subtraction (Prado & Knops, 2024). Similarly, it could be argued that for conceptual understanding of arithmetic principles some children might represent tasks on a mental number line. The extent to which children might employ this strategy might depend on their visuo-spatial WM capacity. While visuo-spatial WM emerged as a significant longitudinal predictor of children's conceptual understanding of the commutative and the inversion principles in Andersson (2010), visuo-spatial WM was not significantly related to concurrent conceptual understanding in Cragg et al. (2017).

Aim of the Current Study

The aim of the current study was to investigate which numerical and domain-general skills are important for the development of conceptual understanding of arithmetic principles. Furthermore, we wanted to compare those predictors to the predictors for two further measures of mathematical performance: speeded arithmetic and mathematical reasoning. We choose speeded arithmetic as a comparator because longitudinal predictors of speeded arithmetic have been researched extensively and this measure draws on procedural skills.

Mathematical reasoning was chosen because, while also measuring children's procedural skills, it measures a wider range of mathematical applied problem solving and also draws on conceptual understanding. In addition there is some evidence for at least partial independence in children's development of speeded arithmetic and mathematical reasoning, because they might rely in part on the development of different white matter brain connections (van Eimeren et al., 2008).

Hypotheses

For the conceptual understanding task, we predicted children would perform best on items requiring an understanding of the commutativity principle and worse on the items testing the understanding of the inverse relationship between addition and subtraction and the understanding of the subtraction complement principle. Furthermore, they would find it easier to provide judgements whether a particular arithmetic problem would help them to solve another arithmetic problem than to provide explanations for why that would be the case.

We expected arithmetic performance, magnitude understanding, multi-digit understanding, as well as non-verbal IQ, verbal and visuo-spatial WM in Year 1 to be significant predictors of conceptual understanding of arithmetic principles two years later in Year 3.

Method

Participants

All children in Year 1 in eleven United Kingdom (U.K.) primary schools were invited to take part in the study; 301 children (mean age = 6 years, 2 months, $SD = 3.7$ months; 141 girls, 160 boys) participated in Year 1. Thirty-one children had a special educational needs (SEN) statement or were receiving support and 14 children were identified as having English as additional language (EAL). We followed up 195 children (96 girls, 99 boys, mean age = 8 years, 2 months, $SD = 3.9$ months, 20 children with a SEN statement, 6 EAL children) when they were in Year 3 (approximately 24 months later). Data for explanations were only available for a subset of children in Year 3 ($N = 138$). A Missing Values Analysis indicated that Little's (1988) test of Missing Completely at Random (MCAR) was not significant, $\chi^2(120) = 123.591$, $p = .393$. Children came from four urban, three town and four rural schools, with a mean deprivation index decile score of 8 (indicating the 30% of least deprived neighbourhoods, [Department for Communities and Local Government, 2015](#)) and an average of 11% of free school meals.

Materials and Procedure

As part of a larger battery, children were assessed on number writing, number reading, number-identification skills, magnitude comparison, arithmetic, working memory, and non-verbal reasoning in Year 1 and on speeded arithmetic, conceptual understanding and mathematical reasoning in Year 3. Magnitude comparison, arithmetic, non-verbal reasoning, speeded arithmetic and mathematical reasoning were administered as paper-and-pencil measures to whole class groups in a fixed order in sessions of 1 hour each. Children were tested individually on number reading, working memory in Year 1, and on conceptual understanding in Year 3. Reliabilities are reported in [Table 1](#).

Year 1 Measures

Number Transcoding

Number Writing – Children were asked to write down a total of 52 items distributed evenly across four subtests. This measure included 4 single-digit numbers, 24 double-digit numbers, 16 three-digit numbers and 8 four-digit numbers. Children were instructed to write Arabic digits to dictation, one item per line, using consecutive lines. Items were

scored correct or incorrect (1, 0) and a total correct score (maximum score = 52) was calculated.

Number Reading – Children were asked to read aloud a list of 52 Arabic numbers, printed in Calibri font, size 20 points, one number per row. The numbers were the same as in the number writing task, but they were ordered by number of digits (increasing in digit string length). A card was used to cover all the numbers and the card was moved down after each number was read allowing the child to see the next number. All children read out all items up to the first four-digit number. If more than three three-digit numbers had been read correctly the child moved on to the remaining four-digit numbers otherwise testing was discontinued. Items were scored correct or incorrect (1, 0) and a total correct score (maximum score = 52) was calculated.

Number Identification – A number identification task from Göbel et al. (2014) with eight items was used to assess children’s ability to identify one-, two- and three-digit Arabic numerals. The experimenter said the target number aloud (e.g. ‘28’), and the children attempted to identify the corresponding Arabic digit strings out of four or five presented response options on the answer sheet. The first item was a single-digit number (‘6’), followed by 4 two-digit numbers (‘14’, ‘28’, ‘52’, and ‘76’) and 3 three-digit numbers (‘163’, ‘235’ and ‘427’). Distractors were chosen on the basis of visual similarity to the target number and common errors with place-value (e.g., for the target number ‘427’, choices were 472, 427, 47, 42, 40027). One point was given per correct response (maximum score = 8).

Number Comparison

Children completed comparison tasks as reported in Göbel et al. (2014). There were six pairs of items on each page, presented in rows. Children were asked to tick the numerically larger item in a pair and to complete as many items as possible (without leaving any pairs out) in 30 seconds. The children were required to wait for the experimenter to say “Go” to begin and to stop as soon as told. The number of correct items per exercise provided an estimate of efficiency in magnitude comparison. There were three symbolic comparison tasks, including one practice task to begin. Pairs of items in the symbolic tasks were single Arabic digits (presented in Calibri font, size 48); in one task the digits were numerically close (with a distance of one or two) and in the other they were numerically far (with a distance of five, six or seven). There were five non-symbolic comparison tasks, including one practice, which was administered first. In the non-symbolic tasks dot arrays were presented; arrays consisted of between 5 and 40 black squares arranged randomly within a 2.5 cm² box on a white background. There were two versions of this task (same size and same surface area). In the same surface area tasks, the total surface area of the squares was matched (in each item pair in order to

prevent discrimination based on visual/physical characteristics rather than magnitude). In the same size tasks array pairs were numerically close or far. In the same surface area tasks, the ratios between the numerosities were 3:4 and 5:6. Further details of the task are provided in Göbel et al. (2014).

Arithmetic

Children completed the Numerical Operations subtest from the Wechsler Individual Achievement Test (WIAT-II UK; Wechsler, 2005) adapted for group use in Year 1. This measure started with six items that involve identifying and writing Arabic digits and counting dots. The remaining nine items were standard arithmetic calculations (addition, subtraction and multiplication) increasing in difficulty. The researcher dictated the first six items and children were allowed 15 minutes to work through the remaining nine items. The number of correct items was scored. Responses to the first six items were excluded from the total score, because they assessed transcoding rather than arithmetic performance. The total correct score for items 7 to 15 provides a conventional measure of arithmetic performance (maximum score = 9).

Non-Verbal IQ

Children completed sets A-C of Raven's Standard Progressive Matrices Plus (Raven et al., 1998) adapted for group use. Following two practice trials completed as a class, children were allowed 15 minutes to complete the remaining test items. The number of correct test items was scored (maximum = 34).

Working Memory

The Forward Digit Recall and Forward Block Recall subtests from the Working Memory Test Battery for Children (Pickering & Gathercole, 2001) were administered individually. Subtests were administered according to the test manual. The number of total test items correct was calculated for each subtest (maximum = 54). The total score for the Forward Digit subtest provided a measure of verbal working memory, and the total score for Block Recall provided a measure of visuo-spatial working memory.

Year 3 Measures

Speeded Arithmetic

In Year 3 children were tested on six speeded arithmetic subtests (addition, addition extra, subtraction, subtraction extra, multiplication and division). Each subtest consisted of three A4 pages. An example calculation item was presented in written format on the first page. The test items (60 items for all addition and subtraction subtests, 56 for the multiplication and division subtests) were printed on Pages 2 and 3 in Calibri font size 24 in two columns per page. For each subtest children were given 60 seconds

to write down the answers to as many calculation items as possible. For the addition, subtraction and multiplication subtests all operands were single digits. For the addition extra, subtraction extra and division subtests the first operand was always a double-digit number and the second operand a single-digit number. One point was given for each correct item. The total number of correct items was calculated separately for each subtest (maximum for each addition/subtraction subtest = 60, maximum for multiplication and division subtests = 56).

Mathematical Reasoning

Children completed an adapted version of the mathematical reasoning subtest from the Wechsler Individual Achievement Test (WIAT-II UK; [Wechsler, 2005](#)). We adapted the subtest for group use by selecting a representative subset of 18 items (item numbers in the original WIAT-II mathematical reasoning test: 10, 15, 16, 19, 22, 23, 25, 26, 29, 30, 31, 32, 34, 35, 36, 37, 42, 48) from the 53 original items and by providing the children with an answer sheet to record their responses. Items consisted of completing number series, arithmetic word problems, interpreting graphs, measurement items, items about time, dates and money and fraction problems. The researcher read each item out aloud and gave children approximately one minute for each item to write down their answer onto the answer sheet. The number of correct answers was scored (maximum = 18).

Conceptual Understanding

In this computerised task we used the conceptual understanding items from [Cragg et al. \(2017\)](#). Children were presented with an addition or subtraction problem with the correct solution followed by another addition or subtraction problem presented without its solution. Children had to decide whether the first problem with its solution could help them to answer the second problem or not. The relationship between the first and the second problem was manipulated and could be: identical (e.g., $23 + 24 = 47$; $23 + 24 = ?$), inverse (e.g., $23 + 24 = 47$; $47 - 23 = ?$), commutative (e.g., $23 + 24 = 47$; $24 + 23 = ?$), subtraction-complement (e.g., $76 - 32 = 44$; $76 - 44 = ?$) or unrelated (e.g., $38 + 23 = 61$; $38 + 32 = ?$). The task was presented in PsychoPy v1.85.3 ([Peirce, 2007](#)) on a 15.6-inch Dell Inspiron 15 laptop (resolution 1920 x 1080 pixels) with an external standard QWERTY keyboard. The screen background was black, numbers and text were presented in white Arial font (size set at 0.15 of screen size).

Judgements – On each trial an addition or subtraction problem with two two-digit operands with its answer was presented in the top line followed by the text ‘If you know that, can it help you solve’ on the next line. This was followed by presentation of an addition or subtraction problem with two two-digit operands without its answer on the following line. Participants pressed the green-stickered button (“L” key) with their right index-finger for a ‘yes’ response and the red-stickered button (“A” key on the left side

of the keyboard) with their left index-finger for a 'no' response. There were five practice trials followed by 24 test trials (6 identical trials, 6 inverse trials, 6 unrelated trials, 2 subtraction-complement trials, and 4 commutativity trials). Accuracy and response times were recorded. The number of correct responses on the test trials was used in the longitudinal models (maximum = 24).

Explanations — For a subset of children ($N = 138$, 68 female, mean age = 8 years 3 months, $SD = 3.9$ months, 15 children with a SEN statement, 4 EAL children), we asked them to provide an explanation of their decision for eight of the 24 items (two identical, two inverse, two unrelated, one commutative and one subtraction-complement, see [Appendix, Table A1](#)) and scored these explanations as correct or incorrect.

Data Analyses

Structural Equation Modelling

The structural equation model of these data (including data from all 301 children in Year 1) was estimated with Mplus Version 8.11 ([Muthén & Muthén, 1998-2011](#)). Missing values were handled with full-information maximum-likelihood estimation. We ran a latent variable path model in which variations in speeded arithmetic, mathematical reasoning, conceptual understanding (correct number of trials for a) judgements, b) explanations) in Year 3 were predicted from constructs measured in Year 1 (arithmetic, magnitude-comparison, number transcoding; visuo-spatial WM; verbal WM; nonverbal abilities; and age).

In this model, in Year 1 all six nonsymbolic and symbolic numerical magnitude-judgment tasks load on a single latent magnitude-comparison factor, whereas the three number transcoding tasks load on a different factor. Preliminary analyses showed that this two-factor model (where the correlation between the independent latent factors is $r = .568$) fitted the data significantly better than a model equivalent to a one-factor solution in which the correlation between the magnitude-comparison and number transcoding factors was fixed to 1, Wald $\chi^2(1) = 101.082$, $p < .001$. Furthermore, when considering the symbolic (digit) and nonsymbolic magnitude-comparison tasks alone, there was no difference between a one-factor magnitude model and a nested model in which the symbolic and nonsymbolic comparison tasks loaded on one factor each, Wald $\chi^2(1) = 2.358$, $p = .125$. Because visuo-spatial and verbal WM, nonverbal ability, arithmetic and mathematical reasoning and conceptual understanding were each assessed by only one indicator, to avoid distortions caused by measurement error, we prespecified the error variance for these measures based on their estimated reliabilities measured by Cronbach's alpha.

In addition, we ran a second longitudinal model using exactly the same data and methods with one exception: in this second model we did not include children's arithmetic performance in Year 1. This model was added to investigate whether the predictive

pattern changes in a model without the auto-regressor. In previous studies arithmetic in earlier years was often a strong predictor of arithmetic in later years. Several of our domain-specific (magnitude comparison, transcoding) and domain-general (verbal and visuo-spatial WM) predictors are typically correlated with arithmetic. Thus, it is possible that some of those predictors do not explain enough unique variance in our outcome measures when arithmetic in Year 1 is included, but they might be significant predictors in a model that does not include arithmetic in Year 1.

Results

The data are available on OSF (see Göbel, 2025S).

Conceptual Understanding

Judgements

On average, in 80.00% ($SD = 15.8\%$) of all trials children judged correctly whether the first item could help them to solve the second item. After we excluded all trials with RTs longer than 10 seconds (236 trials, 6.27%), we calculated mean RTs by type and overall by participants (see Table 1). Children took on average 4362 ms ($SD = 1116$ ms) for correct decisions.

Table 1

Conceptual Understanding: Mean RT in ms and Mean Accuracy in % for Correct Judgements by Item Type

Item type	Judgements			
	N	Accuracy (SD)	N	RT (SD)
Identical	195	94.62 (14.69)	141	3728 (1174)
Inverse	195	64.62 (36.04)	141	4573 (1437)
Commutative	195	94.10 (15.14)	141	4222 (1432)
Subtraction-complement	195	69.20 (38.02)	141	4515 (1631)
Unrelated	195	75.47 (27.72)	141	4772 (1462)

Item type significantly affected accuracy, $F(4, 776) = 59.123$, $p < .001$, $\eta_p^2 = .234$. Children were most accurate in their judgements for identical and commutative items (the difference in accuracy between those two item types was not significant, $p = .66$). Their accuracy on those two item types was significantly higher than their accuracy on all other item types (all $ps < .001$). Accuracy was lowest, and significantly lower than for all other item types, for inverse items (all $ps < .05$).

These effects were mirrored in the reaction time data. Because of empty cells only data from 141 children was included in the statistical analyses of reaction times by item type. Item type significantly affected response times, $F(4, 560) = 22.677$, $p < .001$, $\eta_p^2 = .139$. Children's judgements were significantly faster on identical than any other item types (all $ps < .001$). Children responded also significantly faster to commutative than inverse, subtraction-complement and unrelated items (all $ps < .034$). Children took the longest to respond to unrelated items, but there were no significant differences in RTs between inverse, subtraction-complement and unrelated items.

Explanations

Overall, children gave a correct explanation in 75.72% ($SD = 21.13\%$) of the probed items (see Table 2). The number of correct explanations was strongly correlated with the overall number of correct judgements of whether the solution to the first equation would help with solving the second ($r = 0.808$, $p < .001$, $N = 138$). As for the correct judgements, the percentage of correct explanations varied significantly between item types ($F(4, 576) = 34.583$, $p < .001$, $\eta_p^2 = .194$). Explanations for both identical and commutative items were both significantly more accurate than for any other item types (all $ps < .001$), but their accuracy did not differ significantly between them. The percentage of accurate answers was significantly lower for subtraction-complement and inverse than for all other item types (all $ps < .004$), but their accuracy did not differ significantly from each other. The accuracy for unrelated items sat in the middle and was also significantly different from all other item types (all $ps < .004$).

Table 2

Conceptual Understanding: Mean Accuracy (SD) in % for Correct Judgements and for Correct Explanations, by Item Type, Only for Items With Explanations

Item type	N	Judgements	Explanations
Identical	138	93.12 (20.21)	89.86 (24.29)
Inverse	138	68.84 (41.06)	63.04 (43.96)
Commutative	138	98.55 (11.99)	94.20 (23.45)
Subtraction-complement	138	66.67 (47.31)	58.70 (49.42)
Unrelated	138	78.99 (32.99)	73.55 (36.34)

Comparing Judgements and Explanations

Overall children were significantly more accurate in their judgements than in their explanations ($F(1, 137) = 37.154$, $p < .001$, $\eta_p^2 = .213$). Item type significantly affected accuracy ($F(4, 548) = 30.70$, $p < .001$, $\eta_p^2 = .183$), but the interaction between task and item type was not significant ($F(4, 548) = 1.21$, $p = .308$, $\eta_p^2 = .009$).

Longitudinal Models

Descriptives of all observed variables are shown in Table 3 and correlations between all observed variables are shown in Table 4.

Table 3

Mean Score and Reliability for All Variables in Year 1 and Year 3

Measure	<i>N</i>	<i>M (SD)</i>	Reliability
Year 1			
Number transcoding			
Number writing	301	34.29 (8.41)	.938 ^a
Number identification	301	5.33 (1.56)	.670 ^a
Number reading	301	37.94 (10.30)	.940 ^a
Number comparison			
Digits: far	301	16.02 (6.02)	.653 ^b
Digits: close	301	14.46 (5.69)	.653 ^b
Non-symbolic same size: far	301	19.64 (6.38)	.699 ^b
Non-symbolic same size: close	301	11.17 (4.56)	.699 ^b
3:4 surface area	301	10.73 (4.58)	.666 ^b
5:6 surface area	301	9.28 (4.05)	.666 ^b
Non-verbal IQ	300	14.52 (4.49)	.765 ^a
Verbal working memory	301	25.65 (4.06)	.874 ^a
Visuo-spatial working memory	301	20.08 (4.25)	.859 ^a
Arithmetic	301	4.11 (1.98)	.670 ^a
Year 3			
Speeded arithmetic			
Addition	195	17.24 (6.53)	.917 ^a
Addition Extra	195	10.95 (5.33)	.897 ^a
Subtraction	194	13.32 (5.80)	.913 ^a
Subtraction Extra	194	9.73 (5.60)	.906 ^a
Multiplication	195	8.51 (6.32)	.926 ^a
Division	195	7.47 (6.28)	.925 ^a
Mathematical reasoning	191	12.73 (2.93)	.751 ^a
Conceptual understanding			
Correct judgements	195	19.21 (3.81)	.796 ^a
Correct explanations	138	6.06 (1.69)	.660 ^a

Note. Standard deviations are given in parentheses.

^aCronbach's alpha. ^bparallel test reliability.

Table 4
Correlations Between Observed Variables in the Longitudinal Model

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
Year 1																								
1 Age	-	.277**	.223**	.275**	.273**	.280**	.277**	.301**	.267**	.246**	.280**	.023	.094	.198**	.144*	.167*	.213**	.186**	.164*	.157*	.220**	.172*	.071	
2 Number writing		-	.752**	.818**	.474**	.496**	.455**	.371**	.460**	.381**	.395**	.281**	.350**	.644**	.604**	.657**	.679**	.662**	.586**	.644**	.713**	.421**	.398**	
3 Number identification			-	.651**	.393**	.387**	.337**	.285**	.348**	.290**	.287**	.231**	.277**	.494**	.539**	.553**	.579**	.543**	.472**	.497**	.621**	.590**	.376**	
4 Number reading				-	.421**	.459**	.414**	.347**	.414**	.384**	.370**	.288**	.348**	.588**	.522**	.605**	.625**	.587**	.499**	.531**	.654**	.309**	.305**	
5 NC digit far					-	.654**	.648**	.610**	.514**	.474**	.277**	.131*	.328**	.428**	.422**	.396**	.504**	.425**	.414**	.403**	.437**	.290**	.214**	
6 NC digit close						-	.744**	.706**	.684**	.704**	.267**	.099	.327**	.432**	.447**	.430**	.493**	.475**	.423**	.413**	.458**	.272**	.258**	
7 NC NS SS far							-	.691**	.639**	.666**	.210**	.120*	.265**	.437**	.421**	.366**	.428**	.390**	.349**	.346**	.442**	.324**	.220**	
8 NC NS SS close								-	.627**	.648**	.199**	.047	.214**	.338**	.359**	.309**	.304**	.291**	.323*	.298**	.368**	.233**	.186*	
9 NC NS 3/4									-	.657**	.278**	.109	.259**	.443**	.437**	.446**	.447**	.419**	.411**	.355**	.457**	.327**	.242**	
10 NC NS 5/6										-	.152**	.059	.224**	.376**	.362**	.348**	.373**	.314**	.298**	.272**	.390**	.276**	.200**	
11 Non-verbal IQ											-	.242**	.349**	.363**	.337**	.367**	.366**	.383**	.187**	.274**	.389**	.380**	.353**	
12 Verbal WM												-	.232**	.287**	.095	.119*	.162*	.167*	.087	.141*	.356**	.170*	.198*	
13 Visuo-spatial WM													-	.339**	.366**	.356**	.353**	.266**	.237**	.251**	.335**	.265**	.279**	
14 Arithmetic														-	.484**	.583**	.581**	.551**	.523**	.575**	.645**	.424**	.302**	
Year 3																								
15 Addition															-	.845**	.794**	.747**	.657**	.569**	.650**	.443**	.402**	
16 Addition Extra																-	.818**	.780**	.665**	.644**	.644**	.461**	.410**	
17 Subtraction																	-	.879**	.702**	.719**	.664**	.461**	.398**	
18 Subtraction extra																		-	.691**	.683**	.631**	.437**	.358**	
19 Multiplication																			-	.848**	.604**	.407**	.360**	
20 Division																				-	.605**	.446**	.432**	
21 Mathematical reasoning																					-	.518**	.435**	
22 CU judgements																						-	.816**	
23 CU explanations																							-	

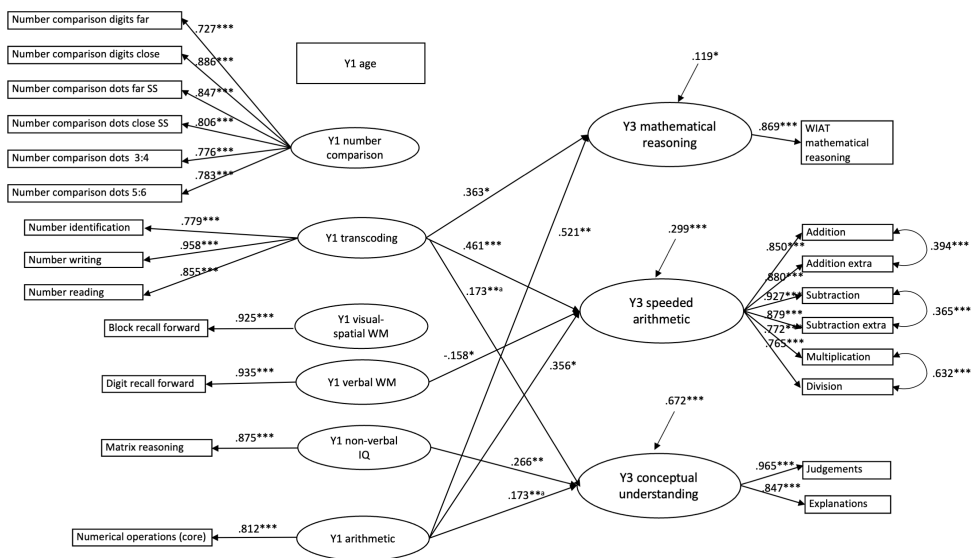
Note. NC = number comparison; CU = conceptual understanding.

Longitudinal Model 1: Including Arithmetic as Predictor in Year 1

Figure 1 shows a latent variable path model in which variations in the conceptual understanding task in Year 3, as well as variations in speeded arithmetic and mathematical reasoning in Year 3, were predicted from all constructs measured in Year 1 (speeded arithmetic, number comparison, transcoding, and visuo-spatial WM, verbal WM, nonverbal abilities; and age). Because the correlation between the independent latent factors transcoding and arithmetic in Year 1 was very high ($r = .827$), to avoid collinearity we fixed their regression paths to be equal for conceptual understanding.

Figure 1

Longitudinal Model 1



Note. Conceptual understanding, speeded arithmetic and mathematical reasoning in Year 3 predicted by all constructs, including arithmetic, measured in Year 1. Ellipses reflect latent variables and rectangles reflect observed variables. Values on one-headed arrows from the latent to the observed variables reflect factor loadings in the measurement model, and values associated with one-headed arrows between latent variables reflect true-score regressions between constructs. The one-headed arrow from the number into the latent variable reflects the residual of the construct. All predictor constructs were correlated (see Table A2), but for clarity these correlations are not shown in the diagram. Solid lines indicate statistically significant relationships.

^aThese two regression paths have been fixed to be equal to avoid collinearity.

Asterisks indicate significant paths (* $p < .05$. ** $p < .01$. *** $p < .001$).

The model fit for this model was not significantly worse than the model without fixed regression paths, Wald $\chi^2(1) = 1.347$, $p = .246$. Thus we are reporting the results for the model with the fixed regression paths here. This model fitted the data very well, $\chi^2(168) = 261.225$, $p < .05$, root-mean-square error of approximation (RSMEA) = .036 (90% CI [.024, .046]), comparative fit index (CFI) = .982, Tucker-Lewis Index (TLI) = .976, standardized root mean residual (SRMR) = .036, which confirms that the factor structure specified in the measurement model for the Year 1 measures was satisfactory. The correlations and regression coefficients between latent variables in this model are shown in the [Appendix, Table A2](#).

Conceptual Understanding — This model shows that there were 3 unique predictors of individual differences in the conceptual understanding task in Year 3: arithmetic ability, number transcoding and non-verbal IQ in Year 1. Conceptual understanding in Year 3 was also significantly correlated with mathematical reasoning ($r = 0.482$, $p = .003$) and speeded arithmetic ($r = 0.238$, $p = .007$) in Year 3. Overall, the model explained 32.8% of the variance in the conceptual understanding task.

Speeded Arithmetic — There were 3 unique predictors of individual differences in speeded arithmetic in Year 3: arithmetic ability, number transcoding and verbal WM in Year 1. Verbal WM in Year 1 was negatively related to speeded arithmetic in Year 3. Speeded arithmetic in Year 3 was also significantly correlated with mathematical reasoning ($r = 0.547$, $p = .001$). Overall, the model explained 70.1% of the variance of speeded arithmetic in Year 3.

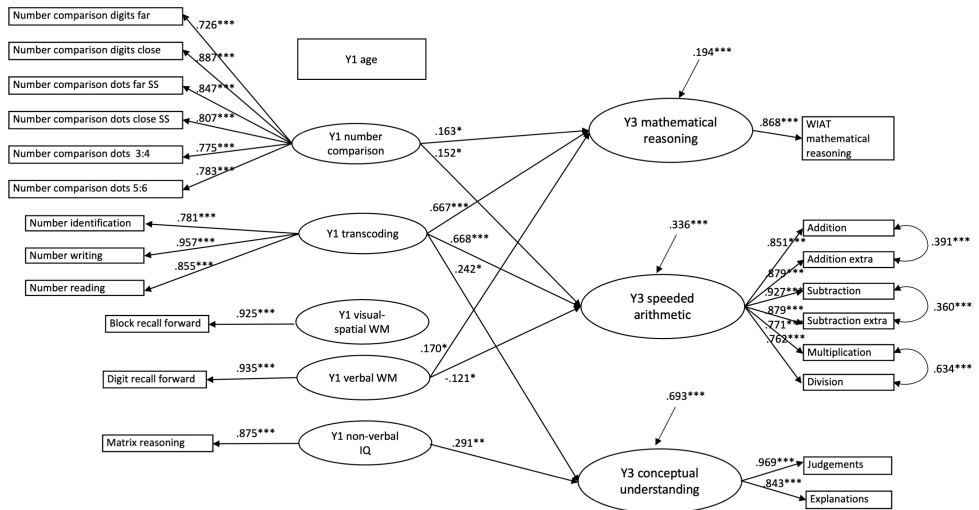
Mathematical Reasoning — Two unique predictors of individual differences in mathematical reasoning in Year 3 emerged: arithmetic ability and number transcoding in Year 1. Overall, the model explained 88.1% of the variance of mathematical reasoning.

Longitudinal Model 2: Without Arithmetic as Predictor in Year 1

When Year 1 arithmetic was dropped as a predictor in our second longitudinal model, the model fit was still good ($\chi^2(175) = 237.673$, $p < .001$, RSMEA = .034 (90% CI [.022, .045]), CFI = .984, TLI = .978, SRMR = .035). In this model (see [Figure 2](#)) transcoding was still a significant longitudinal predictor of conceptual understanding, speeded arithmetic and mathematical reasoning. Now, number comparison emerged as a significant longitudinal predictor of speeded arithmetic and mathematical reasoning, but not of conceptual understanding, in Year 3. In this model without arithmetic as a predictor in Year 1, verbal WM was a significant longitudinal predictor of speeded arithmetic and mathematical reasoning. Visuo-spatial WM was not a significant longitudinal predictor of any of the outcome measures in Year 3.

Figure 2

Longitudinal Model 2



Note. Conceptual understanding, speeded arithmetic and mathematical reasoning in Year 3 predicted by all constructs, measured in Year 1. Ellipses reflect latent variables and rectangles reflect observed variables. Values on one-headed arrows from the latent to the observed variables reflect factor loadings in the measurement model, and values associated with one-headed arrows between latent variables reflect true-score regressions between constructs. The one-headed arrow from the number into the latent variable reflects the residual of the construct. All predictor constructs were correlated (see Table A3), but for clarity these correlations are not shown in the diagram. Solid lines indicate statistically significant relationships. Asterisks indicate significant paths. (* $p < .05$. ** $p < .01$. *** $p < .001$).

Conceptual Understanding – There were two unique predictors of individual differences in the conceptual understanding task in Year 3: number transcoding and non-verbal IQ in Year 1. Conceptual understanding in Year 3 was also significantly correlated with mathematical reasoning ($r = 0.471, p < .001$) and speeded arithmetic ($r = 0.275, p = .001$) in Year 3. Overall, the model explained 30.7% of the variance in the conceptual understanding task.

Speeded Arithmetic – There were three unique predictors of individual differences in speeded arithmetic in Year 3: number transcoding, number comparison and verbal WM in Year 1. Speeded arithmetic in Year 3 was also significantly correlated with mathematical reasoning ($r = 0.604, p < .001$). Overall, the model explained 66.4% of the variance of speeded arithmetic in Year 3.

Mathematical Reasoning – Three unique predictors of individual differences in mathematical reasoning in Year 3 emerged: number transcoding, number comparison and

verbal WM in Year 1. Overall, the model explained 80.6% of the variance of mathematical reasoning.

Discussion

Using a conceptual understanding of arithmetic principles task adapted from [Cragg et al. \(2017\)](#) we found clear inter-individual differences in conceptual understanding in 8 to 9 year-old children as well as evidence for better conceptual understanding of arithmetic principles related to addition than to subtraction. Children's arithmetic skills, number transcoding abilities and their non-verbal cognitive skills in Year 1 predicted over 32% of the variance in conceptual understanding of arithmetic principles two years later. Number transcoding and non-verbal cognitive skills also remained the only significant longitudinal predictors of conceptual understanding in Year 3 in a model without arithmetic performance included as predictor in Year 1.

Commutativity, Subtraction Complement and Inversion

In line with previous findings ([Andersson, 2010](#); [Canobi, 2004](#); [Gilmore & Papadatou-Pastou, 2009](#); [Robinson & Dubé, 2009b](#)) our results show clear inter-individual differences in conceptual understanding of arithmetic principles in 8- to 9-year old children. We also found consistent patterns of differences between the level of understanding for the three principles we investigated. Children were significantly faster and more accurate in their judgements on items testing commutativity than on items testing their understanding of the inversion or subtraction complement principle. Their high accuracy on commutativity items (94% of judgements and explanation were correct) and comparatively low accuracy (below 70%) for inversion and subtraction complement items supports the idea that children first acquire an understanding of additive relationships before they move onto understanding the relationship between addition and subtraction, and subtraction complements (e.g. [Canobi, 2004, 2005](#)).

Our results also provide further support that inversion understanding, at least with two-term problems, emerges relatively late ([Canobi, 2005](#)), because the accuracy of children's judgements on the inverse items was well below ceiling and significantly lower than for all other item types. As predicted and reported before (e.g. [Canobi et al., 2003](#)), children's explicit understanding of arithmetic principles (measured through explanations) was lower than their implicit understanding of the same arithmetic principles.

While we recorded whether children thought that a previous problem could help them to solve the next problem and for a subset of the items also their explanations for why they thought that was, we did not explicitly test whether children actually used the arithmetic principles we tested in the conceptual understanding task spontaneously

when solving arithmetic problems. [Bisanz and LeFevre \(1992\)](#) proposed three possible developmental sequences: 1. Evaluation before application: children might be able to judge whether an arithmetic principle can be applied, before they can accurately apply it themselves, 2. application before evaluation: they might only be able to provide an accurate judgement after they have learnt to apply the principle in their own problem solving, 3. reciprocal development: a bidirectional influence between children's development in their ability to evaluate and to apply arithmetic principles. Our data does not directly speak to this issue, but our longitudinal results provide some tentative evidence for either application before evaluation or reciprocal development.

Longitudinal Predictors of Conceptual Understanding

Children's arithmetic skill in Year 1 emerged as a significant predictor of all three arithmetic outcome measures (speeded arithmetic, mathematical reasoning and conceptual understanding of arithmetic principles) two years later. While our design does not allow us to claim a purely unidirectional influence of earlier procedural arithmetic skill on later conceptual understanding of arithmetic principles (because we have no measure of conceptual understanding in Year 1), our results nevertheless provide evidence against a purely unidirectional influence of earlier conceptual understanding on later procedural skill. As suggested by [Canobi et al. \(2003\)](#) children who develop good procedural arithmetic skills earlier might be able to pay more attention to mathematical relationships which in turn might lead to better conceptual understanding. Individual differences in conceptual understanding in Year 3 were also significantly related to individual differences in speeded arithmetic in Year 3, possibly providing more support for a bidirectional relationship ([Rittle-Johnson et al., 2015](#)).

Interestingly, while the conceptual understanding task tested only arithmetic principles related to addition and subtraction, performance on the conceptual task was significantly correlated with children's concurrent performance on speeded arithmetic which included multiplication and division. Could this partly be because children with better conceptual understanding, in particular on the inverse relationship between addition and subtraction, might also have a better understanding of the inverse relationship between multiplication and division? We did not test for this in our study and existing research on the relationship between addition and subtraction inversion problems and multiplication and division problems suggests that this is unlikely. Several studies in older children found that using the inversion shortcut for addition/subtraction problems did not predict using the shortcut for multiplication/division problems (e.g., [Robinson et al., 2006](#)) and that even after several weeks of training with inversion problems only about a third of 11-year old children used multiplicative and division inversion ([Robinson & Dubé, 2009a](#)).

Our measure of multi-digit understanding, number transcoding, in Year 1 also emerged as a significant predictor of all three arithmetic outcome measures (speeded

arithmetic, mathematical reasoning and conceptual understanding of arithmetic principles) two years later. Interestingly, this was the case whether arithmetic performance in Year 1 was included as a longitudinal predictor in the model or not. Number transcoding has previously been reported as a significant predictor of speeded arithmetic (Göbel et al., 2014; Malone et al., 2021) and in a recent study (Bakker et al., 2024) number reading in preschool emerged as a significant predictor of high achievement in mathematics (top 15%) in Year 1 and Year 3 of primary school. To our knowledge this is the first time that early abilities of multi-digit number understanding have been shown to be a longitudinal predictor of conceptual understanding and mathematical reasoning above and beyond early arithmetic abilities. This is providing further evidence that early understanding of the multi-digit number system and how it maps onto number words is influencing conceptual and further arithmetic development beyond the first year of formal schooling.

An understanding of place value is important for arithmetic understanding (Bisanz & LeFevre, 1992), particularly for multi-digit problems which we used in our conceptual understanding task and also for arithmetic problems that cross decade boundaries and involve carrying or borrowing (Dresen et al., 2020). The importance of understanding the multi-digit number system, here measured as number transcoding, for conceptual understanding is likely to increase over the first few years of formal schooling, when the focus of children's instruction moves onto larger multi-digit numbers.

Children succeeding in our number transcoding task possess a basic understanding of place-value. However, place value is a broader construct and including additional and more explicit measures of place-value understanding (such as asking children to state the number of units, decades and hundreds in a 3-digit number) might be a fruitful approach in future studies, to assess the relative contributions of various aspects of place-value understanding to conceptual understanding.

Furthermore, place-value understanding is clearly also important for other arithmetic principles that we did not assess here, for example decomposition. The use of decomposition strategies (e.g., solving $6+5$ by calculating $6+4+1$) increases in frequency in the first school years (Siegler, 1987). Canobi et al. (1998) investigated addition strategies in 6- to 8 year-old children and found that while most children noticed items where the order of addends had been changed (commutativity), they often failed to notice when addends had been decomposed or recombined. Good place-value understanding is related to the concurrent use of decomposition strategies in arithmetic (Laski et al., 2014). Thus, by including decomposition and recombination items in future assessments of conceptual understanding of arithmetic principles we could compare how (and possibly when) place-value understanding supports the development of conceptual understanding across a range of specific arithmetic principles.

Interestingly, non-verbal IQ uniquely predicted conceptual understanding and was not a significant predictor of speeded arithmetic or mathematical reasoning in Year 3. The relationship between general cognitive skills and conceptual understanding is cur-

rently under-researched. One interesting question is whether this relationship emerged in our study because of the measure of non-verbal reasoning we used. Morsanyi (2020), for example, has suggested that reasoning is an important component of learning new concepts. Alternatively, children with better general cognitive skills might just be faster learners in general and possibly, due to having more cognitive resources, have more capacity earlier on to focus on mathematical relationships. This is supported by a longitudinal study (Andersson, 2010) finding that a composite measure of verbal and non-verbal IQ significantly predicted children's later conceptual understanding of arithmetic principles. Further research is needed to disentangle the influence of verbal and non-verbal reasoning on the development of conceptual understanding of arithmetic principles.

In contrast to some previous studies (Andersson, 2010; Cragg et al., 2017) neither verbal nor visuo-spatial WM in Year 1 significantly predicted conceptual understanding in Year 3, even when we did not include arithmetic as a predictor in Year 1. This was surprising. One possible explanation for visuo-spatial WM not emerging as a predictor could be its concurrent correlation with non-verbal IQ ($r = .35$). While performance on both visuo-spatial WM and non-verbal reasoning in Year 1 were significantly correlated with conceptual understanding, the correlation of conceptual understanding with non-verbal IQ was stronger ($r = .38$ vs $r = .26$). The correlation between verbal WM in Year 1 and conceptual understanding two years later was significant, but small ($r = .20$). The conclusion of our results is that while verbal WM might influence performance on a conceptual understanding task concurrently like in Cragg et al. (2017), it does not seem to predict future conceptual understanding above and beyond procedural arithmetic skill and non-verbal IQ.

In line with other studies (Peng et al., 2016) verbal WM was a significant predictor of speeded arithmetic in Year 3 in both of our longitudinal models providing further evidence that arithmetic facts might be stored in a verbal format (De Smedt & Boets, 2010). Verbal WM only emerged as a significant predictor of mathematical reasoning in Year 3 when we dropped arithmetic as a predictor in Year 1.

This suggests that while good verbal WM in Year 1 was helpful for solving the arithmetic tasks, both in the arithmetic performance measure in Year 1 and in the mathematical reasoning measure, arithmetic performance in Year 1 was a better predictor of later performance in mathematical reasoning than verbal WM. Thus, in the longitudinal model including arithmetic as a predictor verbal WM in Year 1 did not explain any significant variance in mathematical reasoning in Year 3 above and beyond shared variance with arithmetic performance in Year 1. Another explanation for the absence of an association between the working memory measures and conceptual understanding could be due to the nature of the tasks used. For both verbal and visuo-spatial WM measures, we only used forward measures. These measures might be assessing short-term memory rather than working memory and their load on working memory might have been too low for

strong longitudinal effects, especially for later conceptual understanding. While we have tried to investigate visuo-spatial and verbal WM as potential predictors of later conceptual understanding, we did not include other measures of executive function that have been suggested to be involved in conceptual understanding. Several authors suggested that executive functions might be involved when conceptual understanding gets activated (Barrouillet et al., 2004; Unsworth & Engle, 2007). However, Cragg et al. (2017) did not find a significant concurrent relationship between the executive functions of shifting and inhibition and conceptual understanding. They argue that they might not have found a relationship between executive function skills and conceptual understanding because their participants applied already existing understanding and because inhibiting irrelevant information and approaches as well as restructuring and reshaping problems might be only important when conceptual relationships are acquired. Following their argument, we would expect a longitudinal relationship between early EF skills and the development of conceptual understanding. Thus, measuring a wider range of executive functions skills than we did in our current study might be a fruitful avenue for future research on conceptual understanding of arithmetic principles.

Magnitude understanding measured as number comparison was not a significant predictor of any of the arithmetic outcome measures when arithmetic performance in Year 1 was included as a predictor. However, when arithmetic performance in Year 1 was not included as longitudinal predictor, magnitude understanding emerged as a significant predictor for both speeded arithmetic and mathematical reasoning in Year 3. This is in line with previous findings. On the one hand, significant longitudinal relationships between non-symbolic and symbolic number processing and arithmetic (e.g., see meta-analysis by Schneider et al., 2017) have been reported when arithmetic performance at Time 1 has not been controlled for. On the other hand, several longitudinal studies including number transcoding as additional predictor showed that magnitude understanding at the beginning of primary school is no longer a longitudinal predictor of arithmetic when number transcoding and arithmetic performance are included as predictors (e.g., Göbel et al., 2014).

However, magnitude understanding measured as number comparison was not a significant predictor of conceptual understanding whether arithmetic was included as a predictor in Year 1 or not. While magnitude understanding might be essential for an early understanding of addition is more, subtraction is less and thus predict later performance in speeded arithmetic and mathematical reasoning, magnitude understanding might only be tangentially relevant for the arithmetic principles we tested. There is no need to activate magnitude in items testing commutativity. While it could be used for subtraction complement ($c - a = b$ AND $c - b = a$) or inversion ($a + b = c$ AND $c - a = b$) because in both cases the largest number does not change its place, at least for inverse problems children might have reverted to a non-magnitude based strategy, for example that the first operand in a subtraction problem does not change its place.

While our longitudinal model explained a significant amount of variability in children's conceptual understanding in Year 3, the amount of explained variance (Model 1: 32.8%, Model 2: 30.7%) was much smaller than the variance explained by our model for speeded arithmetic (Model 1: 70.1%, Model 2: 66.4%) and mathematical reasoning (Model 1: 88.1%, Model 2: 80.6%). There are at least two possible explanations. First, our measure of conceptual understanding might have been less sensitive. Indeed, in contrast to the measures for speeded arithmetic and mathematical reasoning, participants can in principle get correct responses for the judgements in the conceptual understanding task by guessing. This is unlikely to fully account for the differences in explained variance, because the number of correct explanations provided by our participants was highly correlated with the number of correct judgements ($r = .816$), thus excluding a complete guessing approach for many children. Thus, it is likely that despite including a larger number of cognitive measures in Year 1 than previous studies, there are other important foundations for conceptual understanding of arithmetic principles (such as reasoning and abstraction skills, metacognitive factors and executive skills) that we did not capture in the current study.

Conclusion

In sum, our longitudinal study provides evidence that conceptual understanding of principles related to addition develops before conceptual understanding of principles related to subtraction and shows that children with good early arithmetic and number transcoding skills show better conceptual understanding of arithmetic principles several years later.

Funding: This work was supported by the Economic and Social Research Council [ES/N014677/1, ES/W002914/1] and the FWF, Austria [I 2778-G16]. This research was partially supported by the Research council of Norway (project number 331640) through its Centre of Excellence Scheme.

Acknowledgments: We would like to thank Francina Clayton, Nicoleta Gavrila, Clare Copper, Erin Dysart, Marta Wesierska, Christina Roberts, Rebecca Reed, Philippa Gibbons, Mariela Rios Diaz, Lea Prange, for help with data collection and entry.

Competing Interests: The authors have declared that no competing interests exist.

Author Contributions: *Silke M. Göbel:* Conceptualization; Formal analysis; Funding acquisition; Methodology; Project administration; Supervision; Roles/Writing – original draft, review & editing. *Karin Landerl:* Conceptualization; Funding acquisition; Methodology; Project administration; Supervision; Roles/Writing – review & editing. *Arne O. Lervåg:* Conceptualization; Formal analysis; Methodology; Roles/Writing – original draft, review & editing.

Data Availability: The data are available on OSF (see Göbel, 2025S).

Supplementary Materials

The Supplementary Materials contain the research data for this study (see Göbel, 2025S).

Index of Supplementary Materials

Göbel, S. M. (2025S). *Longitudinal predictors of conceptual understanding of arithmetic principles in Year 3* [Research data and codebook]. OSF. <https://osf.io/8c9bz/>

References

- Andersson, U. (2010). Skill development in different components of arithmetic and basic cognitive functions: Findings from a 3-year longitudinal study of children with different types of learning difficulties. *Journal of Educational Psychology, 102*(1), 115–134. <https://doi.org/10.1037/a0016838>
- Bakker, M., Torbeyns, J., Verschaffel, L., & De Smedt, B. (2024). Cognitive characteristics of children with high mathematics achievement before they start formal schooling. *Child Development, 95*(6), 2062–2081. <https://doi.org/10.1111/cdev.14140>
- Banfi, C., Clayton, F. J., Steiner, A. F., Finke, S., Kemény, F., Landerl, K., & Göbel, S. M. (2022). Transcoding counts: Longitudinal contribution of number writing to arithmetic in different languages. *Journal of Experimental Child Psychology, 223*, Article 105482. <https://doi.org/10.1016/j.jecp.2022.105482>
- Baroody, A. J. (1999). Children's relational knowledge of addition and subtraction. *Cognition and Instruction, 17*(2), 137–175. <https://doi.org/10.1207/S1532690XCI170201>
- Barrouillet, P., Camos, V., Perruchet, P., & Seron, X. (2004). ADAPT: A developmental, asemantic, and procedural model for transcoding from verbal to Arabic numerals. *Psychological Review, 111*(2), 368–394. <https://doi.org/10.1037/0033-295X.111.2.368>
- Bisanz, J., & LeFevre, J.-A. (1992). Understanding elementary mathematics. In J. I. D. Campbell (Ed.), *Advances in Psychology: Vol. 91. The nature and origins of mathematical skills* (pp. 113–136). Elsevier. [https://doi.org/10.1016/S0166-4115\(08\)60885-7](https://doi.org/10.1016/S0166-4115(08)60885-7)
- Braithwaite, D. W., & Sprague, L. (2021). Conceptual knowledge, procedural knowledge, and metacognition in routine and nonroutine problem solving. *Cognitive Science, 45*(10), Article e13048. <https://doi.org/10.1111/cogs.13048>
- Canobi, K. H. (2004). Individual differences in children's addition and subtraction knowledge. *Cognitive Development, 19*(1), 81–93. <https://doi.org/10.1016/j.cogdev.2003.10.001>
- Canobi, K. H. (2005). Children's profiles of addition and subtraction understanding. *Journal of Experimental Child Psychology, 92*(3), 220–246. <https://doi.org/10.1016/j.jecp.2005.06.001>
- Canobi, K. H. (2009). Concept-procedure interactions in children's addition and subtraction. *Journal of Experimental Child Psychology, 102*(2), 131–149. <https://doi.org/10.1016/j.jecp.2008.07.008>

- Canobi, K. H., Reeve, R. A., & Pattison, P. E. (1998). The role of conceptual understanding in children's addition problem solving. *Developmental Psychology, 34*(5), 882–891. <https://doi.org/10.1037/0012-1649.34.5.882>
- Canobi, K. H., Reeve, R., & Pattison, P. E. (2003). Patterns of knowledge in children's addition. *Developmental Psychology, 39*(3), 521–534. <https://doi.org/10.1037/0012-1649.39.3.521>
- 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>
- Crooks, N. M., & Alibali, M. W. (2014). Defining and measuring conceptual knowledge in mathematics. *Developmental Review, 34*(4), 344–377. <https://doi.org/10.1016/j.dr.2014.10.001>
- Dehaene, S. (2001). Précis of The Number Sense. *Mind & Language, 16*(1), 16–36. <https://doi.org/10.1111/1468-0017.00154>
- de Hevia, M. D., Castaldi, E., Streri, A., Eger, E., & Izard, V. (2017). Perceiving numerosity from birth. *The Behavioral and Brain Sciences, 40*, Article e169. <https://doi.org/10.1017/S0140525X16002090>
- Department for Communities and Local Government. (2015). *The English indices of deprivation 2015 statistical release*. <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015>
- De Smedt, B., & Boets, B. (2010). Phonological processing and arithmetic fact retrieval: Evidence from developmental dyslexia. *Neuropsychologia, 48*(14), 3973–3981. <https://doi.org/10.1016/j.neuropsychologia.2010.10.018>
- Dresen, V., Pixner, S., & Moeller, K. (2020). Effects of place-value and magnitude processing on word problem solving. *Cognitive Development, 54*, Article 100876. <https://doi.org/10.1016/j.cogdev.2020.100876>
- Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Harvard University Press.
- Gilmore, C. K. (2023). Understanding the complexities of mathematical cognition: A multi-level framework. *The Quarterly Journal of Experimental Psychology, 76*(9), 1953–1972. <https://doi.org/10.1177/17470218231175325>
- Gilmore, C. K., & Papadatou-Pastou, M. (2009). Patterns of individual differences in conceptual understanding and arithmetical skill: A meta-analysis. *Mathematical Thinking and Learning, 11*(1-2), 25–40. <https://doi.org/10.1080/10986060802583923>
- Gilmore, C. K., & Spelke, E. S. (2008). Children's understanding of the relationship between addition and subtraction. *Cognition, 107*(3), 932–945. <https://doi.org/10.1016/j.cognition.2007.12.007>
- 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>
- Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the “Number Sense”: The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457–1465. <https://doi.org/10.1037/a0012682>

- Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1–27). Lawrence Erlbaum Associates.
- Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., & Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *Journal of Experimental Child Psychology, 116*(1), 45–58. <https://doi.org/10.1016/j.jecp.2013.02.001>
- Laski, E. V., Ermakova, A., & Vasilyeva, M. (2014). Early use of decomposition for addition and its relation to base-10 knowledge. *Journal of Applied Developmental Psychology, 35*(5), 444–454. <https://doi.org/10.1016/j.appdev.2014.07.002>
- Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*(6), 1292–1300. <https://doi.org/10.1111/j.1467-7687.2011.01080.x>
- Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>
- Malone, S. A., Pritchard, V. E., & Hulme, C. (2021). Separable effects of the approximate number system, symbolic number knowledge, and number ordering ability on early arithmetic development. *Journal of Experimental Child Psychology, 208*, Article 105120. <https://doi.org/10.1016/j.jecp.2021.105120>
- McCrink, K., & Wynn, K. (2004). Large-number addition and subtraction by 9-month-old infants. *Psychological Science, 15*(11), 776–781. <https://doi.org/10.1111/j.0956-7976.2004.00755.x>
- McNeil, N. M., Hornburg, C. B., Devlin, B. L., Carrazza, C., & McKeever, M. O. (2019). Consequences of individual differences in children’s formal understanding of mathematical equivalence. *Child Development, 90*(3), 940–956. <https://doi.org/10.1111/cdev.12948>
- Morsanyi, K. (2020). Reasoning skills in individuals with mathematics difficulties. In *Handbook of educational psychology and students with special needs* (pp. 510–533). Routledge. <https://doi.org/10.4324/9781315100654-24>
- Muthén, L. K., & Muthén, B. O. (1998–2011). *Mplus user's guide* (6th ed.). Los Angeles, CA, USA: Muthén & Muthén.
- Peirce, J. W. (2007). PsychoPy – Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8–13. <https://doi.org/10.1016/j.jneumeth.2006.11.017>
- Peng, P., Namkung, J., Barnes, M., & Sun, C. (2016). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *Journal of Educational Psychology, 108*(4), 455–473. <https://doi.org/10.1037/edu0000079>
- Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychological Science, 24*(6), 1037–1043. <https://doi.org/10.1177/0956797612464057>
- Pickering, S., & Gathercole, S. E. (2001). *Working memory test battery for children (WMTB-C)*. London, United Kingdom: Psychological Corporation.

- Prado, J., & Knops, A. (2024). Spatial attention in mental arithmetic: A literature review and meta-analysis. *Psychonomic Bulletin & Review*, 31(5), 2036–2057.
<https://doi.org/10.3758/s13423-024-02499-z>
- Rasmussen, C., Ho, E., & Bisanz, J. (2003). Use of the mathematical principle of inversion in young children. *Journal of Experimental Child Psychology*, 85(2), 89–102.
[https://doi.org/10.1016/S0022-0965\(03\)00031-6](https://doi.org/10.1016/S0022-0965(03)00031-6)
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven's standard progressive matrices and vocabulary scales*. Oxford Psychologists Press.
- Rittle-Johnson, B., Schneider, M., & Star, J. R. (2015). Not a one-way street: Bidirectional relations between procedural and conceptual knowledge of mathematics. *Educational Psychology Review*, 27(4), 587–597. <https://doi.org/10.1007/s10648-015-9302-x>
- Robinson, K. M., & Dubé, A. K. (2009a). A microgenetic study of the multiplication and division inversion concept. *Revue Canadienne de Psychologie Experimentale [Canadian Journal of Experimental Psychology]*, 63(3), 193–200. <https://doi.org/10.1037/a0013908>
- Robinson, K. M., & Dubé, A. K. (2009b). Children's understanding of addition and subtraction concepts. *Journal of Experimental Child Psychology*, 103(4), 532–545.
<https://doi.org/10.1016/j.jecp.2008.12.002>
- Robinson, K. M., Ninowski, J. E., & Gray, M. L. (2006). Children's understanding of the arithmetic concepts of inversion and associativity. *Journal of Experimental Child Psychology*, 94(4), 349–362. <https://doi.org/10.1016/j.jecp.2006.03.004>
- Sarnecka, B. W., & Wright, C. E. (2013). The idea of an exact number: Children's understanding of cardinality and equinumerosity. *Cognitive Science*, 37(8), 1493–1506.
<https://doi.org/10.1111/cogs.12043>
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science*, 20(3), Article e12372.
<https://doi.org/10.1111/desc.12372>
- Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, 116(3), 250–264.
<https://doi.org/10.1037/0096-3445.116.3.250>
- Siegler, R. S., & Lortie-Forgues, H. (2015). Conceptual knowledge of fraction arithmetic. *Journal of Educational Psychology*, 107(3), 909–918. <https://doi.org/10.1037/edu0000025>
- Sophian, C. (1995). Representation and reasoning in early numerical development: Counting, conservation, and comparisons between sets. *Child Development*, 66(2), 559–577.
<https://doi.org/10.2307/1131597>
- Steiner, A. F., Finke, S., Clayton, F. J., Banfi, C., Kemény, F., Göbel, S. M., & Landerl, K. (2021). Language effects in early development of number writing and reading. *Journal of Numerical Cognition*, 7(3), 368–387. <https://doi.org/10.5964/jnc.6929>

- Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038–1066. <https://doi.org/10.1037/0033-2909.133.6.1038>
- van Eimeren, L., Niogi, S. N., McCandliss, B. D., Holloway, I. D., & Ansari, D. (2008). White matter microstructures underlying mathematical abilities in children. *Neuroreport*, *19*(11), 1117–1121. <https://doi.org/10.1097/WNR.0b013e328307f5c1>
- Wechsler, D. (2005). *Wechsler individual achievement test – Second UK edition (WIAT-IIUK)*. Harcourt Assessment.

Appendix

Table A1

List of Items With Template Explanations in the Conceptual Understanding Task

Item	Item type	Correct answer	Correct explanation
63 - 31 = 32 63 - 31 =	Identical	yes	The numbers are the same
23 + 24 = 47 32 + 24 =	Unrelated	no	The numbers are different
63 - 31 = 32 63 - 32 =	Subtraction - complement	yes	The answer has been swapped around
23 + 24 = 47 23 + 24 =	Identical	yes	The numbers are the same
63 - 31 = 32 63 - 13 =	Unrelated	no	The numbers are different
23 + 24 = 47 24 + 23 =	Commutativity	yes	The answer is the same when the numbers are in different order
23 + 24 = 47 47 - 23 =	Inverse	yes	It's the opposite sum
63 - 31 = 32 31 + 32 =	Inverse	yes	It's the opposite sum

Table A2

Correlations Between Independent Latent Factors (in Black), Regressions Coefficients Between Latent Variables (in Black Bold and Italic) and the Residual Correlations Between the Latent Outcomes (in Italic) in Longitudinal Model 1 (With Arithmetic Included as Predictor)

Latent Variables	1	2	3	4	5	6	7	8	9
1. Y1 number comparison	–	.568***	.361***	.123	.323***	.616***	.081	.058	.090
2. Y1 transcoding		–	.401***	.323***	.469***	.823***	.461***	.363*	.173**
3. Y1 visuo-spatial WM			–	.269***	.430***	.450***	.049	-.030	.032
4. Y1 verbal WM				–	.297***	.377***	-.158*	.114	-.005
5. Y1 non-verbal IQ					–	.514***	.081	.046	.266**
6. Y1 arithmetic						–	.356*	.521**	.173**
7. Y3 speeded arithmetic							–	.547**	.238*
8. Y3 mathematical reasoning								–	.482**
9. Y3 conceptual understanding									–

Table A3

Correlations Between Independent Latent Factors (in Black), Regressions Coefficients Between Latent Variables (in Black Bold and Italic) and the Residual Correlations Between the Latent Outcomes (in Italic) in the Longitudinal Model 2 (Without Arithmetic Included as Predictor)

Latent Variables	1	2	3	4	5	6	7	8
1. Y1 number comparison	–	.568***	.361***	.123	.323***	.152*	.163*	.134
2. Y1 transcoding		–	.402***	.323***	.471***	.668***	.667***	.242*
3. Y1 visuo-spatial WM			–	.269***	.430***	.082	.019	.052
4. Y1 verbal WM				–	.297***	-.121*	.170*	.019
5. Y1 non-verbal IQ					–	.127	.112	.291**
6. Y3 speeded arithmetic						–	.604***	.275***
7. Y3 mathematical reasoning							–	.471***
8. Y3 conceptual understanding								–

<div style="text-align: center;">  <p>Journal of Numerical Cognition (JNC) is the official journal of the Mathematical Cognition and Learning Society (MCLS).</p> </div>	<div style="text-align: center;">  <p>PsychOpen GOLD is a publishing service provided by the Leibniz Institute for Psychology (ZPID), Germany.</p> </div>
--	--