Journal of Experimental Child Psychology 223 (2022) 105482

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Transcoding counts: Longitudinal contribution of number writing to arithmetic in different languages

Chiara Banfi [a,b,*], Francina J. Clayton [c], Anna F. Steiner [a,d], Sabrina Finke [a], Ferenc Kemény [a,e], Karin Landerl [a,f,g], Silke M. Göbel [c,h]

[a] Institute of Psychology, University of Graz, 8010 Graz, Austria
[b] Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, 8036 Graz, Austria
[c] Department of Psychology, University of York, York YO10 5DD, UK
[d] FH JOANNEUM, University of Applied Sciences, 8020 Graz, Austria
[e] Institute of Education and Psychology at Szombathely, Eötvös Loránd University, 1075 Budapest, Hungary
[f] BioTechMed-Graz, 8010 Graz, Austria
[g] Department of Cognitive Science, Macquarie University, Sydney, New South Wales 2109, Australia
[h] Department of Special Needs Education, University of Oslo, 0318 Oslo, Norway

A R T I C L E I N F O

Article history:
Received 9 December 2021
Revised 24 May 2022
Available online 1 July 2022

Keywords:
Arithmetic
Coss-linguistic comparison
Longitudinal study
Number processing
Number writing
Transcoding

A B S T R A C T

Number writing involves transcoding from number words (e.g., “thirty-two”) to written digit strings (32) and is an important unique predictor of arithmetic. The existing longitudinal evidence about the relation between transcoding and arithmetic is mostly language specific. In languages with number word inversion (e.g., German), the order of tens and units is transposed in spoken number words compared with Arabic numbers. This makes transcoding more challenging than in languages without number word inversion (e.g., English). In the current study, we aimed to understand whether the contribution of number writing to the development of arithmetic is similar in languages with and without number word inversion. German-speaking children (n = 166) and English-speaking children (n = 201) were followed over the first 3 years of primary school. In a series of multiple linear regressions, we tested whether number writing of multi-digit numbers was a significant unique predictor of arithmetic after controlling for well-known non-numerical predictors (nonverbal reasoning and working memory) and numerical predictors (symbolic and

* Corresponding author at: Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, 8036 Graz, Austria.
E-mail address: chiara.banfi1@gmail.com (C. Banfi).

https://doi.org/10.1016/j.jecp.2022.105482
0022-0965/

nonsymbolic magnitude comparison). Number writing in Grade 1 predicted arithmetic in Grades 1, 2, and 3 over and above the other predictors. Crucially, number writing performance was of comparable importance for arithmetic development in German- and English-speaking children. Our findings extend previous evidence by showing that transcoding predicts the development of arithmetic skills during the first 3 years of primary school in languages with and without number word inversion.

Introduction

Number writing relies on transcoding spoken number words (e.g., "thirty-two") to written Arabic digits (e.g., 32). According to the triple code model (Dehaene, 1992), there are three distinct representational codes for numbers. Number words (e.g., "four") are processed by the auditory–verbal word frame, grounded on language skills. Arabic digits (e.g., 4) are handled by the visual Arabic number form. The analog magnitude representation is thought to process the meaning of numbers in a nonsymbolic approximate form. Transcoding enables shifting between symbolic codes, in the case of number writing, from the auditory–verbal word frame (input code) to the visual Arabic number form (output code). The ability to integrate these two number codes supports incremental understanding of the number system (Yuan, Prather, Mix, & Smith, 2019). Grasping the number system (number system knowledge) refers to the understanding of how numbers are represented and starts with single-digit knowledge (Nuerk, Moeller, Klein, Willmes, & Fischer, 2011; von Aster & Shalev, 2007): Digits from 0 to 9 are the basic building blocks of any Arabic single- and multi-digit construction, similarly to letters for reading, as pointed out by Clayton et al. (2020) and Malone, Burgoyne, & Hulme, 2020. Number system knowledge further requires understanding the positional system. This is also known as the place-value principle, the notion that the magnitude of each single digit depends on its position in the digit string (e.g., 2 in 213 represents hundreds, 2 in 123 represents decades). Transcoding supports the consolidation of place-value understanding. In line with this notion, a recent study by Cheung and Ansari (2021) showed that number reading accuracy, a form of transcoding, predicted the ability to compare multi-digit numbers, including decade-unit-incongruent number pairs (e.g., 32 vs. 27, where 3 > 2 and 2 < 7), a type of comparison that requires grasping the place-value notational system. We and others have argued that the ability to learn multi-digit Arabic numeral symbols and to match them with their corresponding spoken number words is a foundational skill for later arithmetic development (Clayton et al., 2020; Göbel, Watson, Lervåg, & Hulme, 2014; Habermann, Donlan, Göbel, & Hulme, 2020; Malone et al., 2020; Steiner, Banfi, et al., 2021).

Cross-linguistic differences in number transcoding

Transcoding performance is susceptible to the lexical and morphological specificities of the verbal number system. For example, the lexical primitive corresponding to the Arabic number 2 is "two" in English and "zwei" in German. Transcoding procedures enable us to arrange these lexical units according to a structure. Certain procedures, such as additive rules, are common in most, if not all, number languages ("four hundred [and] fifty-six" is transcoded as 400 + 50 + 6 = 400506, which results in 456 by over-writing the zeros). Some transcoding procedures, however, are language specific. Some languages, for instance, apply an inversion rule, which requires reversing the order of digits in the decade and unit positions in spoken number words compared with their order in the written Arabic digit string (46 would be pronounced "six-and-forty"). Given these language-specific differences in transcoding, in the current study we set out to investigate whether the contribution of number transcoding to arithmetic development is modified by language. To this aim, we compared a language with number word inversion (German) with a language without number word inversion (English).

The presence of the inversion rule affects transcoding; between 40% and 50% of all transcoding errors committed by German-speaking children in first grade are inversion errors (Steiner, Finke, et al., 2021; Zuber, Pixner, Moeller, & Nuerk, 2009). In contrast, in English-speaking first graders (i.e., in children using a language without number word inversion), inversion errors make up less than 10% of all transcoding errors (Steiner, Finke, et al., 2021). Similar findings indicating a higher proportion of inversion errors in languages with versus without number word inversion come from Dutch–French (Imbo, Vanden Bulcke, De Brauwer, & Fias, 2014), German–Japanese comparisons (Moeller, Zuber, Olsen, Nuerk, & Willmes, 2015), and Czech, which allows both inverted and noninverted word forms (Pixner et al., 2011).

The highest proportion of inversion errors is usually observed during the first years of primary school, the period when children are becoming increasingly familiar with multi-digit numbers. Cross-sectional findings indicate that inversion errors diminish with increasing grades; interestingly, however, they do not completely disappear by the end of primary school (van der Ven, Klaiber, & van der Maas, 2017). Moreover, experiments focusing on cross-format integration of spoken number words and Arabic digits showed that number word inversion incurs a cost in terms of reaction times not only in German-speaking children (Poncin, Van Rinsveld, & Schiltz, 2020; Steiner, Banfi, et al., 2021) but even in German-speaking adults (Steiner, Banfi, et al., 2021).

Number transcoding and arithmetic

The notion that transcoding is fundamental for the development of arithmetic skills has so far been confirmed in a number of studies, showing both unique concurrent (Clayton et al., 2020; Imbo et al., 2014; Sowinski et al., 2015; van der Ven et al., 2017) and longitudinal (Göbel et al., 2014; Habermann et al., 2020; Malone et al., 2020; Moeller, Pixner, Zuber, Kaufmann, & Nuerk, 2011) contributions of transcoding (including measures of number writing) to arithmetic over and above other numerical and non-numerical predictors.

This evidence, however, does not speak directly to the question of whether the influence of number writing is comparable across languages. Most studies selectively focused on either languages with number word inversion (Imbo et al., 2014; Moeller et al., 2011; van der Ven et al., 2017) or languages without number word inversion (Göbel et al., 2014; Habermann et al., 2020; Malone et al., 2020; Sowinski et al., 2015). Those studies consistently indicated a reliable contribution of transcoding to arithmetic for languages both with and without number word inversion. However, it remains unclear whether this contribution is of comparable strength. The presence of the inversion rule constitutes an additional challenge for children learning an inverted number language (Pixner et al., 2011; Steiner, Finke, et al., 2021; Zuber et al., 2009). Thus, it might be expected that the relation between transcoding and arithmetic is stronger in languages with versus without number word inversion due to the additional complexity added by the inversion rule. Cross-linguistic concurrent evidence so far has not corroborated this hypothesis. Indeed, Clayton et al. (2020) found a comparable contribution of writing double-digit numbers to arithmetic in first graders from a language background with (German) and without (English) number word inversion. In a similar vein, Steiner, Banfi, et al. (2021) performed an auditory–visual number matching task with two- and three-digit numbers and reported a positive concurrent association between arithmetic and the speed with which both German- and English-speaking second and third graders identified the correct match. This association was of roughly similar size in third graders. Clayton et al. (2020) argued that the association between transcoding and arithmetic is not driven by the higher processing demands of the inverted number language. Thus, these findings speak for a comparable importance of multi-digit transcoding for arithmetic in languages both with and without number word inversion. Cross-linguistic longitudinal evidence so far is not available; longitudinal studies to date have focused on only one language at a time (Göbel et al., 2014; Habermann et al., 2020; Malone et al., 2020; Moeller et al., 2011). Therefore, it is unclear whether transcoding contributes to later arithmetic performance to a comparable extent in languages with and without number word inversion. The current study aimed to fill this gap by investigating the predictive role of transcoding for later arithmetic longitudinally in a cross-linguistic design, comparing two languages, one with and one without number word inversion.

Magnitude processing and arithmetic

Arithmetic development is supported by a number of numerical skills (Lyons, Price, Vaessen, Blomert, & Ansari, 2014) and non-numerical skills (Cragg, Keeble, Richardson, Roome, & Gilmore, 2017). Thus, the important question in the context of the current study is whether transcoding contributes to arithmetic after controlling for other important skills. One of the most studied and also highly discussed numerical predictors of arithmetic is magnitude processing. Magnitudes can be presented in a nonsymbolic format as dot arrays or in a symbolic format as digits. Nonsymbolic magnitude processing refers to the ability to compare or estimate numerosities in an approximate fashion by accessing their analog representations, as theorized in the triple code model (Dehaene, 1992). According to Dehaene (2009), nonsymbolic magnitude processing provides the foundation for later symbolic number processing with Arabic digits (including arithmetic) by mapping approximate quantities to exact symbolic representations. In line with this theory, meta-analyses reported reliable moderate associations between performance on approximate number processing tasks and performance on mathematics (Chen & Li, 2014; Fazio, Bailey, Thompson, & Siegler, 2014; Schneider et al., 2017). Notably, these associations are stronger in children younger than 6 years (Fazio et al., 2014), suggesting that approximate number processing might be particularly relevant when children start to learn symbolic number forms.

Interestingly, in a meta-analysis, Schneider et al. (2017) showed significantly higher associations between arithmetic and symbolic magnitude processing than between arithmetic and nonsymbolic magnitude processing. Arithmetic strongly relies on the understanding and manipulation of numbers in the symbolic format (i.e., as verbal and Arabic representations). Thus, format similarity may explain the stronger association between arithmetic and symbolic processing than between arithmetic and nonsymbolic processing.

A number of longitudinal studies (Göbel et al., 2014; Habermann et al., 2020; Malone et al., 2020) that investigated the contribution of transcoding to arithmetic included magnitude comparison as a numerical predictor. Findings from these studies are mixed due to methodological and age differences. The first longitudinal study addressing this research question by Göbel et al. (2014) followed 165 English-speaking (UK) children from Grade 1 to Grade 2 and found that, in addition to early arithmetic, performance in a transcoding task was the only additional predictor of later arithmetic. The composite factor of magnitude comparison (including nonsymbolic and symbolic tasks) made no relevant contribution to arithmetic. A subsequent study assessed 71 younger English-speaking (UK) children from 4 to 6 years of age (Habermann et al., 2020) and again found that a composite score of transcoding (including number identification, number writing, and number reading) was the only unique predictor of later arithmetic competence in Grade 1. Magnitude processing did not predict arithmetic after accounting for transcoding in the model. Although these findings are in line with Göbel et al. (2014), it should be noted that the study by Haberman et al. (2020) was based on a rather small sample and therefore it might not have had enough statistical power to detect the contribution of magnitude processing measures. A study by Malone et al. (2020) was based on a much larger sample of 519 English-speaking children assessed at the beginning of the preparation year (4–6 years of age) and again 1 year later. Their findings indicate that Arabic number knowledge (a composite factor of number writing and number identification) made a reliable longitudinal contribution to arithmetic 12 months later. In contrast to Göbel et al. (2014) and Habermann et al. (2020), nonsymbolic magnitude comparison was also a significant predictor of later arithmetic, although with a smaller relative contribution compared with number knowledge. In line with previous evidence (Fazio et al., 2014), the results by Malone et al. (2020) indicate that nonsymbolic magnitude processing may play a prominent role for arithmetic before or at the very start of primary school, when children begin to be formally taught about the symbolic number system. Note, however, that symbolic magnitude comparison was not considered in the study by Malone et al. (2020), and therefore it remains unclear whether it would have made a reliable contribution to arithmetic.

Working memory

Working memory is highly relevant for arithmetic. Calculation requires temporary storage and retrieval of factual knowledge and task-relevant information as well as manipulation of quantities during arithmetic procedures (Cragg & Gilmore, 2014; Cragg et al., 2017). Transcoding procedures, on the other hand, build on working memory resources (e.g., Barrouillet, Camos, Perruchet, & Seron, 2004). Storage and manipulation of verbal and visuospatial information were found to play a significant role in number writing (Clayton et al., 2020; Steiner, Banfi, et al., 2021; Zuber et al., 2009). Therefore, it is important to account for this non-numerical dimension in any model investigating the unique contribution of transcoding to arithmetic.

Working memory is a multidimensional construct (Baddeley, 2012; Baddeley & Hitch, 1974) involving verbal and visuospatial temporary storage as well as manipulation via the so-called central executive. These three aspects of working memory were assessed in the current study, and we expected a reliable contribution of the central executive to arithmetic, as previously shown (Caviola, Colling, Mammarella, & Szűcs, 2020; Clayton et al., 2020; De Vita, Costa, Tomasetto, & Passolunghi, 2022; Meyer, Salimpoor, Wu, Geary, & Menon, 2010; Soltanlou, Pixner, & Nuerk, 2015; Van de Weijer-Bergsma, Kroesbergen, & Van Luit, 2015). Predictions related to the verbal and visuospatial temporary storage components are less straightforward because the available evidence is mixed due to different study designs and task formats (De Vita et al., 2022; Meyer et al., 2010; Soltanlou et al., 2015; Van de Weijer-Bergsma et al., 2015).

The current study

The primary research question of the current study was whether number writing assessed in Grade 1 predicts arithmetic longitudinally in Grades 2 and 3—over and above other numerical and non-numerical predictors—to a similar extent in children from a language background with (German) and without (English) number word inversion.

Our cross-linguistic design enabled us to test which of two predictions based on previous literature fits the data best. On the one hand, we might expect a stronger contribution of transcoding to arithmetic in the language with number word inversion compared with the language without number word inversion due to the higher transcoding demands related to the inversion rule. On the other hand, based on previous concurrent findings (Clayton et al., 2020), we might anticipate a comparable contribution of number writing to arithmetic in the two languages. The latter finding would corroborate the notion that transcoding has a similar cross-linguistic importance because it supports the consolidation of number system knowledge (including Arabic number knowledge and place-value understanding), a sine qua non prerequisite for arithmetic competence in any language irrespective of the presence or absence of number word inversion.

An innovative feature of the current study is that we used a transcoding measure that includes a large variety of multi-digit numbers (two-, three-, and four-digit structures), thereby tapping into advanced transcoding skills in a cross-linguistic setting. This is a critical feature compared with previous studies that either assessed young children and thus mostly focused on one- and two-digit number transcoding only (Habermann et al., 2020; Malone et al., 2020) or investigated more complex transcoding ability but in only one language (Dietrich, Huber, Dackermann, Moeller, & Fischer, 2016; Imbo et al., 2014; Moeller et al., 2011; Moura et al., 2013, 2015; Sowinski et al., 2015; van der Ven et al., 2017).

The second goal of the current study was to disentangle the contributions of symbolic and nonsymbolic magnitude comparison to arithmetic in a design that also includes number transcoding as a predictor. Based on previous evidence of decreasing importance of nonsymbolic skills for arithmetic with increasing proficiency in symbolic processing (Fazio et al., 2014), we expected to find a stronger contribution of symbolic magnitude comparison than nonsymbolic magnitude comparison to arithmetic, especially by the time children develop more advanced symbolic number competence in Grades 2 and 3. However, as already shown and argued elsewhere (e.g., Göbel et al., 2014; Habermann et al., 2020), it is also possible that symbolic magnitude processing will not emerge as a reliable predictor of arithmetic over and above number writing. This finding could indicate that the ability to transcode between spoken number words and Arabic digits exceeds the more basic ability to understand Arabic digits. On the one hand, multi-digit transcoding requires advanced symbolic number knowledge extending beyond single digits. On the other hand, transcoding involves verbal–visual transfer, which is not necessarily required during magnitude comparison tasks with single digits.

Non-numerical predictors include working memory domains and nonverbal reasoning. By adding them in our regression models, we aimed to account for covariance between number writing and arithmetic related to domain-general factors influencing both constructs.

Method

Participants

This study was part of a larger longitudinal cross-linguistic project that followed German- and English-speaking primary school children from Grade 1 to Grade 3 (see also Clayton et al., 2020; Finke et al., 2021; Steiner, Banfi, et al., 2021; Steiner, Finke, et al., 2021). German-speaking children in Graz (Austria) came from a middle-income urban school district. English-speaking children in Yorkshire (UK) 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 neighborhoods) (Department for Communities and Local Government, 2015) and an average of 11% of free school meals. Longitudinal data from Grade 1 to Grade 3 for the current study variables were available for 166 German-speaking children (47% female and 53% male; age Grade 1: *M* = 7 years 2 months, *SD* = 3 months; age Grade 2: *M* = 8 years 2 months, *SD* = 3 months; age Grade 3: *M* = 9 years 1 month, *SD* = 3 months) and 201 English-speaking children (49% female and 51% male; age Grade 1: *M* = 6 years 3 months, *SD* = 4 months; age Grade 2: *M* = 7 years 3 months, *SD* = 4 months; age Grade 3: *M* = 8 years 3 months, *SD* = 4 months). The two groups were matched on duration of formal education. English children start school 1 year earlier than Austrian children, and therefore there was an age difference of approximately 11 months between the two groups.

The study was conducted in accordance with the principles of the Declaration of Helsinki. The University of Graz and University of York psychology department ethics committees granted ethical approval for the study. Head teachers and parents provided consent for children to take part.

Tasks and stimuli

As part of a larger assessment battery, children completed measures of number writing, arithmetic, symbolic and nonsymbolic magnitude comparison, working memory, nonverbal reasoning at Time 1 (Grade 1), and arithmetic at Time 2 (Grade 2) and Time 3 (Grade 3). Tasks were administered in a fixed order as paper-and-pencil measures to a whole class group in sessions of 1 h each. Children were also tested individually on measures of working memory at Time 1.

Number writing

Children were asked to write to dictation a total of 52 items distributed evenly across four sessions. The items included 4 one-digit numbers, 24 two-digit numbers, 16 three-digit numbers, and 8 four-digit numbers. Items were scored as correct (1) or incorrect (0), and a total correct score (maximum = 52) was calculated. Cronbach's alpha was .96 in the German-speaking sample and .94 in the English-speaking sample.

Arithmetic

Children completed the Numerical Operations subtest from the Wechsler Individual Achievement Test (WIAT-II UK; (Wechsler, 2005)) adapted for group use. At Time 1, this measure started with 6 items that involve identifying and writing Arabic digits and counting dots. The remaining 9 items were standard arithmetic calculations (addition, subtraction, and multiplication) increasing in difficulty. The researcher dictated the first 6 items, and children were allowed 15 min to work through the remaining 9 items. The number of correct calculations was scored. Responses to the first 6 items were excluded from the total score because these items explicitly test number identification, counting, and basic transcoding rather than arithmetic skills; the total correct score for Items 7 to 15 provides a conventional measure of arithmetic performance (Time 1 maximum score = 9). Cronbach's alpha for Time 1 arithmetic was .60 in the German-speaking sample and .70 in the English-speaking sample.

At Time 2, 8 new items were added (including multi-digit subtraction and addition, multiplication, and division). Instructions and task duration were the same as at Time 1. Again, responses to the first 6 items were excluded from the total score; the total correct score for Items 7 to 23 provides a conventional measure of arithmetic performance (Time 2 maximum score = 17). Cronbach's alpha for Time 2 arithmetic was .73 in the German-speaking sample and .78 in the English-speaking sample.

At Time 3, 9 more items were added (including subtraction and addition with multi-digit numbers, decimal numbers and fractions, multiplication, and division). The first 6 items of the Time 1 and Time 2 versions were no longer presented because we expected ceiling effects. Instructions and task duration were the same as at Time 1 (Time 3 maximum score = 26). Cronbach's alpha for Time 3 arithmetic was .75 in the German-speaking sample and .87 in the English-speaking sample.

Magnitude comparison

Children completed symbolic and nonsymbolic magnitude comparison tasks as reported in a previous study (Göbel et al., 2014). Six pairs of items were presented on each page of an A5 booklet (5.83×8.27 inches). Children were instructed to process as many pairs as possible in 30 s by ticking the numerically larger quantity in each pair. Two practice tasks (one with digits and one with dots) and six different tasks were given in the following order: digit comparison far, dot comparison close, digit comparison close, dot comparison far, dot comparison with ratio 5:6, and dot comparison with ratio 3:4. Performance was measured as the number of correctly solved items (maximum total score = 60 for the same-size digit and dot comparison tasks and maximum score = 48 for the two same-surface-area dot comparison tasks). *Symbolic magnitude comparison* was tested in two separate conditions with single-digit numbers from 1 to 9. Conditions differed in numerical distance; whereas one condition consisted of number pairs with a small numerical distance (one or two; digit comparison close) between both digits, the other condition consisted of pairs with a large distance (five, six, or seven; digit comparison far). Parallel test reliability was .77 for the German-speaking sample and .77 for the English-speaking sample. The symbolic magnitude comparison score was calculated by averaging the z-values of the total correct scores of both conditions. *Nonsymbolic magnitude comparison* was tested in four separate conditions with pairs of arrays containing 5 to 40 black squares presented in a box. The format was analogous to the symbolic magnitude comparison but involved dots instead of digits. Children were explicitly instructed not to count. Two conditions included different numbers of equally sized squares, ranging from 5 to 13 per array. In one of these same-size conditions, the numerical distance between the pairs of arrays was small (one or two; dot comparison close). The other same-size condition involved pairs of arrays with a large numerical distance (five, six, or seven; dot comparison far). The remaining two nonsymbolic conditions controlled for surface area of the squares to prevent judgments from being made based on surface area or "blackness" of the arrays. These same-surface-area conditions included arrays of 20 to 40 squares, differing by ratios of 3:4 and 5:6. Parallel test reliability was .75 for the German-speaking sample and .87 for the English-speaking sample. The nonsymbolic magnitude comparison score was calculated by averaging the z-values of the total correct scores across the four conditions.

Working memory

The Forward and Backward Digit Recall and Forward Block Recall subtests from the Working Memory Test Battery for Children (Pickering & Gathercole, 2001) were administered individually. An additional Backward Block Recall task was administered; this task used the same digits from the Forward Block Recall task presented in a different order. In the Forward Digit subtest, children were asked to repeat, in the same order, a verbal sequence of single digits dictated by the researcher. In the Backward Digit subtest, children were asked to repeat the digit sequence in a backward order. In the Forward Block subtest, children were presented with a set of nine identical blocks arranged on a board and were asked to watch the researcher tap a sequence of blocks and then to tap the block sequence in the same order. In the Backward Block task, children were asked to tap the sequence in a backward order. Practice items preceded the administration of test trials. For all working memory tasks, 6 items were included per span length. If children correctly recalled 4 of 6 items of a span, they proceeded directly to the next span and were given credit for any omitted trial (move-on rule). The task was discontinued as soon as there were 3 incorrect items within a certain span (discontinuation rule). The number of correct sequences was calculated for each subtest (maximum = 54 for Digit Forward, 36 for Digit Backward, 54 for Block Forward, and 48 for the Block Backward). The total score of the Forward Digit Recall subtest provided a measure of verbal working memory, and the total score of Forward Block Recall provided a measure of visuospatial working memory. To calculate the measure of central executive function, *z*-scores of Backward Digit and Backward Block Recall were averaged within each language group. For the English-speaking children, Cronbach's alpha was .87 for Forward Digit Recall, .81 for Backward Digit Recall, .86 for Forward Block Recall, and .87 for Backward Block Recall. Item-level data were not available for the German-speaking children.

Nonverbal reasoning

Children completed sets A to C of Raven's Standard Progressive Matrices Plus (Raven, Raven, & Court, 1998) adapted for group use. Following two practice trials completed as a class, children were allowed 15 min to independently work on the remaining test items. The number of correct test items was scored (maximum = 34). Cronbach's alpha was .77 for the German-speaking sample and .75 for the English-speaking sample.

Statistical analysis

Demographic, cognitive, and numerical measures included in the study are descriptively displayed as means and standard deviations separately in each language group. We computed Spearman correlations to investigate the association among all variables. We further performed a series of hierarchical regression analyses to investigate the concurrent and longitudinal contributions of numerical and non-numerical predictors to arithmetic in both language groups. In the concurrent model, the first step included non-numerical predictors. In the second step, we added symbolic and nonsymbolic magnitude processing tasks. The third step tested whether number writing explained additional unique variance in arithmetic. Finally, in the last step, we included the number writing by language group interaction term with the aim of testing whether the regression coefficients of number writing differed significantly in German- and English-speaking children. In a second set of regressions, we investigated the contributions of numerical and non-numerical predictors to Grade 2 and Grade 3 arithmetic. Because we were particularly interested in assessing the specific relation of these predictors with arithmetic longitudinally in Grades 2 and 3, we removed the variance of arithmetic that could be explained by past arithmetic performance. We did so by controlling for the autoregressive effect of arithmetic in the previous grade in the first step of each longitudinal model. For the following steps, we applied the same logic as used in the concurrent regression models and added non-numerical and numerical predictors in successive steps.

Variables included in the regression models were *z*-standardized within each language group to control for baseline differences (intercepts) due to the 1-year age lag. Note that, due to the use of *z*-values, the effect of language group was negligible because variables were forced to have $M = 0$ and $SD = 1$ in each language group. This strategy allowed us to test the interaction term without confounds related to differences in the intercepts of predictors.

Analyses were conducted in R (Version 4.1.0) and IBM SPSS Statistics (Version 28).

Results

Descriptive statistics

Table 1 reports the descriptive statistics of the numerical and cognitive measures separately for each language group. Variables were generally normally distributed with few exceptions. To ensure

Table 1
Descriptives for predictor measures and arithmetic in Grades 1, 2, and 3 by language group.

	German-speaking (n = 166)		English-speaking (n = 201)				
	M	*SD*	*M*	*SD*	*d*	p^a	Age-adjusted p^b
Grade 1 predictor measures							
Nonverbal IQ	16.13	4.69	12.42	4.44	0.81	<.001	.774
Forward Digit Recall	25.74	3.49	25.48	3.84	0.07	.504	.342
Forward Block Recall	22.57	3.21	19.96	4.19	0.69	<.001	.037
Backward Digit Recall	10.08	2.66	8.48	2.90	0.57	<.001	.375
Backward Block Recall	13.46	4.67	10.37	4.51	0.67	<.001	.133
Digit comparison, far	18.55	5.60	16.19	6.11	0.40	<.001	.583
Digit comparison, close	18.35	4.81	14.89	5.63	0.66	<.001	.978
Dot comparison, far	24.46	9.48	19.76	6.36	0.59	<.001	.193
Dot comparison, close	12.51	4.22	11.43	4.34	0.25	.016	.037
Dot comparison, 3:4	12.45	3.79	10.82	4.58	0.38	<.001	.275
Dot comparison, 5:6	10.68	3.25	9.26	4.05	0.38	<.001	.275
Number writing	34.16	10.82	34.85	8.42	−0.07	.507	<.001
Arithmetic							
Grade 1	5.46	1.42	4.09	1.97	0.79	<.001	.382
Grade 2	11.23	2.65	8.82	3.13	0.82	<.001	.184
Grade 3	16.56	3.51	13.69	5.34	0.63	<.001	.864

[a,b]The p values refer to unadjusted (p^a) and age-adjusted (p^b) language group comparisons.

that the results of group comparisons were not affected by a slight deviation from normality, we compared language groups with both nonparametric and parametric tests and obtained very consistent results. Only the parametric tests are presented in Table 1.

The two language groups differed significantly in most measures, with German-speaking children outperforming English-speaking children. Note, however, that these differences may be explained by the 1-year age gap between language groups. Indeed German- and English-speaking children did not differ significantly in the vast majority of dimensions once age was entered in the models as a covariate. There were, however, three exceptions. Forward Block Recall yielded a significant group difference notwithstanding the presence of age as a covariate. The inspection of the estimated marginal means revealed that German-speaking children outperformed English-speaking children after controlling for the age lag (estimated marginal means difference = 1.49, SE = 0.71). In addition, after including age as a covariate, there was a significant difference between the language groups in the nonsymbolic magnitude comparison task with small numerical distance (dot comparison close). The inspection of the estimated marginal means revealed that the pattern of group difference was reversed by the inclusion of age as a covariate: English-speaking children had higher estimated marginal means (M = 12.67, SE = 0.42) than German-speaking children (M = 11.01, SE = 0.49). Number writing performance, which initially did not differ between language groups, showed significant differences after adding age as a covariate. The inspection of the marginal means of the analysis of covariance (ANCOVA) revealed that German-speaking children had lower estimates (M = 30.34, SE = 1.08) than English-speaking children (M = 38.00, SE = 0.93). This finding clearly indicates that after controlling for the between-group age gap, the English-speaking first graders outperformed the German-speaking first graders, most likely due to the presence of the inversion rule in German.

We tested for distance effects in the magnitude comparison tasks, expecting to observe significantly higher accuracy in far conditions than in close conditions, as previously shown (Holloway & Ansari, 2009). For the symbolic comparisons, the distance effect was small and not significant in the German-speaking sample, $t(165)$ = 0.58, p =.562, d = 0.05, and was larger and significant in the English-speaking sample, $t(200)$ = 3.63, p <.001, d = 0.26. Distance effects for nonsymbolic comparisons were significant in both language groups [German-speaking children: $t(165)$ = 19.63, p <.001, d = 1.52; English-speaking children: $t(200)$ = 26.37, p <.001, d = 1.86]. The comparison of the two additional nonsymbolic tasks with different ratios yielded significant distance effects in both language groups [German-speaking children: $t(165) = 6.84$, $p < .001$, $d = 0.53$; English-speaking children: $t(200) = 6.12$, $p < .001$, $d = 0.43$].

Associations among predictors and arithmetic

Correlations among predictors and arithmetic are shown in Table 2 separately for each language group. Numerical and non-numerical predictors showed positive significant relations with arithmetic at Times 1, 2, and 3. Notably, number writing displayed the strongest association with arithmetic in both language groups ($.59 \leq r$s $\leq .67$). The correlation among numerical and non-numerical predictors was weak or in a few cases moderate, indicating absence of collinearity among predictors. One exception is represented by the two magnitude comparison measures, which were strongly correlated in both samples (rs $\geq .69$). Given the debate about differential contributions of the two types of magnitude comparison (Fazio et al., 2014; Schneider et al., 2017), we decided to keep them as separate variables in our analyses.

Concurrent prediction of arithmetic in Grade 1

Results of the concurrent hierarchical regression model are reported in Table 3. Non-numerical predictors added in the first step accounted for 26% of the variance. Age, nonverbal reasoning, and the central executive predicted Grade 1 arithmetic significantly. Symbolic and nonsymbolic magnitude processing were entered in the second step and resulted in a significant change in R^2 ($\Delta R^2_{\text{adjusted}} = .05$, $p < .001$); however, only nonsymbolic magnitude comparison had a significant impact. The inclusion of number writing as a predictor in the third step explained further additional unique variance ($\Delta R^2_{\text{adjusted}} = .14$, $p < .001$). Finally, the interaction between number writing and language group added in the last step resulted in a nonsignificant change in R^2 ($\Delta R^2_{\text{adjusted}} = .002$, $p = .262$).

Longitudinal prediction of arithmetic

Results of the longitudinal regression models are reported in Table 4. In the model predicting arithmetic in Grade 2, the first step including arithmetic measured in Grade 1 accounted for 33% of the variance. Non-numerical predictors were added in the second step and resulted in a significant R^2 change ($\Delta R^2_{\text{adjusted}} = .06$, $p < .001$), with nonverbal reasoning and the central executive being significant predictors. Magnitude comparison included in the third step contributed a small but significant portion of additional unique variance in Grade 2 arithmetic ($\Delta R^2_{\text{adjusted}} = .01$, $p = .016$). Note, however, that only symbolic magnitude comparison was a significant predictor. Number writing entered in the fourth step resulted in a significant change of R^2 ($\Delta R^2_{\text{adjusted}} = .08$, $p < .001$). Finally, the interaction between number writing and language group produced a nonsignificant increase in R^2 ($\Delta R^2_{\text{adjusted}} < .001$, $p = .586$).

We observed a similar overall pattern for the prediction of arithmetic in Grade 3. A substantial portion of variance in Grade 3 arithmetic was explained by arithmetic measured in Grade 2 ($R^2_{\text{adjusted}} = .47$). Non-numerical predictors were added in the second step and resulted in a significant R^2 change ($\Delta R^2_{\text{adjusted}} = .09$, $p < .001$); nonverbal reasoning and the central executive were significant regressors. Magnitude comparison included in the third step contributed a small but significant portion of additional unique variance in Grade 3 arithmetic ($\Delta R^2_{\text{adjusted}} = .01$, $p = .002$). It is important to note that, unlike in the model predicting Grade 2 arithmetic, for Grade 3 arithmetic nonsymbolic magnitude comparison was a significant predictor ($\beta = .17$, $p = .002$), but symbolic magnitude comparison was not. Number writing entered in the fourth step resulted in a significant change in R^2 ($\Delta R^2_{\text{adjusted}} = .03$, $p < .001$), thereby explaining additional unique variance in Grade 3 arithmetic. Finally, the interaction between number writing and language group again produced only a nonsignificant increase in R^2 ($\Delta R^2_{\text{adjusted}} < .001$, $p = .941$).

Equivalence testing

The concurrent and longitudinal regression models consistently revealed a nonsignificant contribution of the interaction between number writing and language. However, the absence of a significant effect does not necessarily imply a null effect. Thus, we conducted an equivalence testing analysis

Table 2
Correlation coefficients (Spearman's rho) for the association between predictors and arithmetic at Times 1, 2, and 3 shown separately for the two language groups.

	Age (months) in Grade 1	Nonverbal reasoning	Verbal working memory	Visuospatial working memory	Central executive	Symbolic magnitude comparison	Nonsymbolic magnitude comparison	Number writing	Arithmetic Grade 1	Arithmetic Grade 2	Arithmetic Grade 3
Age (months) in Grade 1		.31**	.04	.06	.18**	.34**	.36**	.33**	.21**	.22**	.20**
Nonverbal reasoning	.20**		.23**	.32**	.34**	.34**	.29**	.38**	.39**	.37**	.48**
Verbal working memory	−.14	.23**		.17*	.42**	.13	.14*	.34**	.31**	.27**	.28**
Visuospatial working memory	.14	.36**	.15		.35**	.34**	.29**	.32**	.32**	.38**	.37**
Central executive	.13	.38**	.26**	.43**		.31**	.36**	.56**	.52**	.42**	.52**
Symbolic magnitude comparison	−.02	.22**	.08	.36**	.33**		.80**	.56**	.44**	.44**	.45**
Nonsymbolic magnitude comparison	−.02	.22**	.15	.27**	.34**	.69**		.54**	.47**	.40**	.45**
Number writing	.11	.36**	.16*	.21**	.31**	.29**	.38**		.62**	.67**	.67**
Arithmetic Grade 1	.10	.33**	.18*	.17*	.35**	.24**	.28**	.60**		.52**	.60**
Arithmetic Grade 2	.17*	.42**	.18*	.17*	.42**	.24**	.22**	.59**	.58**		.70**
Arithmetic Grade 3	.16*	.51**	.19*	.24**	.42**	.27**	.34**	.62**	.56**	.65**	

Note. English-speaking children are above the diagonal, and German-speaking children are below the diagonal.
* $p < .05$.
** $p < .01$.

Table 3
Hierarchical multiple regression model predicting Grade 1 arithmetic in German- and English-speaking children.

	Model 1		Model 2		Model 3		Model 4	
	β	*p*	*B*	*p*	*β*	*p*	*β*	*p*
Age (months) in Grade 1	.09	.044	.06	.176	.01	.899	.01	.790
Nonverbal reasoning	.21	<.001	.19	<.001	.13	.005	.12	.006
Verbal working memory	.09	.062	.08	.078	.03	.425	.04	.363
Visuospatial working memory	.05	.299	.02	.723	.02	.724	.02	.730
Central executive	.31	<.001	.25	<.001	.13	.007	.14	.005
Symbolic magnitude comparison			.02	.762	–.04	.565	–.03	.655
Nonsymbolic magnitude comparison			.22	.001	.12	.046	.12	.052
Number writing					.48	<.001	.52	<.001
Language by number writing							–.07	.262
F	27.38		24.30		38.73		34.59	
p	<.001		<.001		<.001		<.001	
Adjusted R^2	.26		.31		.45		.45	

Table 4
Hierarchical multiple regression models predicting Grade 2 and Grade 3 arithmetic in German- and English-speaking children.

	Model 1		Model 2		Model 3		Model 4		Model 5	
Model predicting Grade 2 arithmetic	*β*	*p*	*β*	*p*	*β*	*p*	*β*	*p*	*β*	*p*
Arithmetic in Grade 1	.58	<.001	.43	<.001	.41	<.001	.23	<.001	.23	<.001
Age (months) in Grade 1			.06	.179	.05	.257	.01	.796	.01	.851
Nonverbal reasoning			.14	.004	.13	.008	.11	.015	.11	.015
Verbal working memory			.03	.546	.03	.467	.004	.919	.002	.966
Visuospatial working memory			.07	.157	.04	.448	.04	.392	.04	.391
Central executive			.14	.007	.13	.013	.07	.153	.07	.174
Symbolic magnitude comparison					.17	.009	.12	.040	.12	.048
Nonsymbolic magnitude comparison					–.06	.335	–.11	.076	–.11	.080
Number writing							.42	<.001	.39	<.001
Language by number writing									.03	.586
F	184.20		39.57		31.25		39.19		35.23	
p	<.001		<.001		<.001		<.001		<.001	
Adjusted R^2	.33		.39		.40		.48		.48	
Model predicting Grade 3 arithmetic	*β*	*p*	*β*	*p*	*β*	*p*	*β*	*p*	*β*	*p*
Arithmetic in Grade 2	.69	<.001	.51	<.001	.49	<.001	.37	<.001	.37	<.001
Age (months) in Grade 1			–.01	.881	–.02	.574	–.04	.215	–.04	.223
Nonverbal reasoning			.21	<.001	.21	<.001	.20	<.001	.20	<.001
Verbal working memory			.002	.957	–.003	.931	–.02	.521	–.02	.529
Visuospatial working memory			.04	.320	.03	.440	.03	.358	.03	.359
Central executive			.19	<.001	.16	<.001	.12	.004	.12	.004
Symbolic magnitude comparison					–.06	.320	–.06	.223	–.06	.229
Nonsymbolic magnitude comparison					.17	.002	.12	.027	.12	.028
Number writing							.27	<.001	.27	<.001
Language by number writing									-.004	.941
F	324.51		77.51		61.47		63.05		56.59	
p	<.001		<.001		<.001		<.001		<.001	
Adjusted R^2	.47		.56		.57		.60		.60	

of the semi-partial correlations of the interaction effects following the two one-sided tests (TOST) procedure for correlations (Lakens, 2017; Lakens, Scheel, & Isager, 2018). We applied the method in R using the package TOSTER (Lakens, 2017). Note that semi-partial correlations reflect the association between one predictor and the dependent variable, controlling for the influence of the other independent variables on the predictor of interest. Thus, they are particularly suited to evaluate the amount of unique variance in arithmetic that is accounted for by the interaction term.

In a first step, we defined the smallest effect size of interest by means of a sensitivity analysis in G*Power (Faul, Erdfelder, Buchner, & Lang, 2009). Results indicated that, given an alpha level of .05, power of 80%, and N = 367, we would be able to reliably detect a correlation coefficient of .15 with a two-sided test. We further tested whether the 90% confidence interval of the semi-partial correlation coefficients of the interaction terms fell within the interval of equivalence [−.15; .15]. The TOST procedure confirmed that all three semi-partial correlations were included in the interval of equivalence (concurrent model Grade 1: r = −.04, TOST 90% confidence interval (CI) [−.13; .04]; longitudinal model Grade 2: r = .02, TOST 90% CI [−.07; .11]; longitudinal model Grade 3: r = −.002, TOST 90% CI [−.09; .08]). These results indicate that the number writing by language group interaction terms in the three regression models were statistically equivalent to a null effect; that is, all three coefficients were smaller than the smallest meaningful effect.

Discussion

The aim of this study was to directly compare the longitudinal contribution of number transcoding assessed in Grade 1 to arithmetic in Grades 2 and 3 in German- and English-speaking children. In line with previous evidence (Clayton et al., 2020; Göbel et al., 2014; Habermann et al., 2020; Malone et al., 2020), transcoding was an important predictor of arithmetic from Grade 1 to Grade 3 after accounting for other numerical and non-numerical dimensions that are highly relevant for arithmetic. Thus, we provide further evidence that higher transcoding skills are strongly related to higher arithmetic competence later on.

In a previous study (Clayton et al., 2020) based on the same developmental cohort as the current study, we found a reliable and comparable concurrent unique contribution of two-digit number writing to arithmetic in German- and English-speaking Grade 1 children. The current study extends the concurrent findings by Clayton et al. (2020) by using a longitudinal perspective. Here, we provide evidence that number writing measured in Grade 1 is an important predictor of future arithmetic skills assessed in Grades 2 and 3 in languages with and without number word inversion. A second important feature of the current study is that we addressed advanced multi-digit transcoding beyond two-digit numbers (two-, three-, and four-digit number writing). This enabled us to explore the relation between number transcoding and arithmetic cross-linguistically within a broader range of performance. Thus, our results are more easily generalizable to everyday school activities.

Transcoding and arithmetic

Transcoding is a challenging task for children at the start of primary school, especially in languages with number word inversion (Steiner, Finke, et al., 2021). Our results indicate that English-speaking children outperformed German-speaking children in transcoding once the age gap was taken into account. One might expect that number writing, a particular form of transcoding, would reveal a stronger contribution to arithmetic in languages with versus without number word inversion due to the higher complexity of transcoding in languages with number word inversion. The results of the current study speak against this hypothesis. Number writing made a similarly important longitudinal contribution to arithmetic development in German and English; the number of correctly written numbers in Grade 1 was a significant predictor of arithmetic performance throughout the first 3 years of formal schooling. This indicates that the presence of the inversion rule in the German language does not increase the relative importance of transcoding as a predictor of arithmetic development. Our results are consistent with the idea that transcoding performance is associated with the development of number system knowledge of single- and multi-digit numbers independent of language-specific transcoding procedures. This number system knowledge, in turn, supports arithmetic performance. Based on their cross-sectional findings in a sample of 4- to 7-year-olds, Cheung and Ansari (2021) argued that children learn multi-digit numbers through three main developmental steps. In the beginning, children probably recognize multi-digit numbers as a whole and interpret their magnitude based on the length of the digit string. Later on, they learn that a multi-digit number can be split into different parts according to the position of the digits and that the leftmost positions have the highest values.

Finally, they grasp the base-10 principle, which means they understand that each position corresponds to a specific base-10 value (e.g., units, decades, hundreds). Using this framework, we argue that transcoding practice supports the buildup of place-value understanding, and more specifically the understanding of positional knowledge, by training verbal–visual or visual–verbal number format conversion.

In line with the idea that Arabic digit knowledge is necessary but not sufficient to explain the complexity of number system knowledge, symbolic magnitude comparison, a task based on single-digit processing, had a positive and moderate correlation with arithmetic in both language groups. Symbolic magnitude comparison made a unique contribution to variance in Grade 2 arithmetic over and above other non-numerical predictors. However, number writing entered in the next step explained additional unique variance in arithmetic. This finding is in line with the notion that number writing exceeds the more basic ability of single-digit processing because transcoding also requires manipulation of numerals according to the positional system. It should be noted, however, that both the number writing and arithmetic tasks in the current study involved the manipulation of multi-digit numbers, thereby tapping onto similarly complex symbolic processing skills. In contrast, symbolic magnitude comparison, assessed with single digits only, was related to more basic number knowledge and thus might be less relevant for multi-digit arithmetic. This is a limitation of the current study because our transcoding measure differed from the symbolic magnitude comparison task not only in the dimension of interest (i.e., the additional requirement of positional knowledge) but also in the numerical complexity of the stimuli. Future research on number transcoding should include multi-digit comparison measures as a competing predictor of arithmetic.

Our findings with respect to the relation between number transcoding and arithmetic are largely in line with previous longitudinal literature (Göbel et al., 2014; Habermann et al., 2020; Malone et al., 2020). However, Moeller et al. (2011) reported divergent results. They investigated the longitudinal relationship between Grade 1 number transcoding and symbolic magnitude comparison with two-digit number pairs and Grade 3 arithmetic in German-speaking children. Results of their regression models revealed that the error rate in the symbolic comparison task measured in Grade 1 was the only reliable predictor of the error rate in arithmetic in Grade 3. Although the correlation between symbolic comparison in Grade 1 and arithmetic in Grade 3 (r =.25) was of a size very similar to the correlation between transcoding in Grade 1 and arithmetic in Grade 3 (r =.24), there was no significant unique contribution of transcoding to later arithmetic. There are two likely explanations for these findings that are not in line with our results. First, the symbolic magnitude comparison task with two-digit number pairs used by Moeller et al. (2011) was more complex than the symbolic magnitude comparison task we used with single digits. Two-digit comparison requires basic place-value understanding in addition to Arabic digit knowledge. Thus, the comparison and transcoding tasks in Moeller et al.'s study shared more characteristics than those in the current study, and indeed performance on those two tasks was more strongly correlated in Moeller et al.'s study (r =.54) than in our study (r =.29 for the German-speaking sample). This larger overlap might partly explain why symbolic comparison emerged as a unique predictor of arithmetic in their study but not in our study. Second, Moeller et al. (2011) used a computerized reaction time design for both arithmetic performance and symbolic magnitude comparison, but not for transcoding. In both their arithmetic and symbolic number comparison tasks, children were instructed to choose the correct answer from two possible solutions presented on the screen. The higher similarity of task formats between the arithmetic and magnitude comparison tasks compared with the number writing tasks might also explain their regression results. Future studies should investigate in more detail the role of task similarity in order to more clearly differentiate the relative importance of multi-digit transcoding and multi-digit symbolic comparison for arithmetic development.

Magnitude comparison and arithmetic

Our second research question addressed the relative contributions of symbolic and nonsymbolic magnitude comparison to arithmetic development. Nonsymbolic magnitude comparison was a small but significant predictor of concurrent arithmetic in Grade 1, whereas symbolic magnitude comparison did not explain variance in the dependent measure at this time point. This pattern changed in Grade 2, when symbolic but not nonsymbolic magnitude comparison explained significant variance in arithmetic performance. Taken together, these findings are consistent with the idea that nonsymbolic magnitude processing is relevant for early arithmetic skills and decreases in importance through time in favor of symbolic magnitude processing (Fazio et al., 2014) because number processing becomes more formalized and symbol-based. Our results for predicting Grade 3 arithmetic, however, are difficult to reconcile with this view because nonsymbolic magnitude comparison was again a significant predictor of arithmetic in Grade 3, whereas symbolic magnitude comparison was not. There are different explanations for this result that apply to the models in Grades 1 and 2 as well. Magnitude comparison was repeatedly shown to be a construct that changes its structure through grade levels. At preschool age, two distinct components can be identified, namely the nonsymbolic and symbolic ones (Habermann et al., 2020). However, as proficiency in number processing increases through grade levels, these two components become less distinct and a one-dimensional structure of magnitude comparison is found to have a better fit as compared with the two-component structure in school-aged children (Göbel et al., 2014; Habermann et al., 2020). This switch in structure is due to the fact that performances in the symbolic and nonsymbolic subdomains become more strongly associated through time as children become proficient in number processing and can easily map symbolic numerosities on the nonsymbolic ones (and vice versa). Partly related to this argument, symbolic magnitude comparison with single digits may lose explanatory power for arithmetic through grade levels because the complexity of arithmetic tasks with multi-digit numbers exceeds the more basic ability to grasp single digits.

The direction of the causal relation between symbolic and nonsymbolic magnitude processing is currently hotly debated (see, e.g., Goffin & Ansari, 2019) and exceeds the aim of the current study. It is, however, worth mentioning that recent evidence consistently reported a unidirectional causal relation between symbolic and nonsymbolic number processing, opposite to what previous theories had proposed (see, e.g., Piazza et al., 2010). 1); Lau et al. (2021), for instance, found that in young children, from kindergarten to Grade 2, a model with symbolic processing skills predicting later nonsymbolic processing skills had a better fit than a model with nonsymbolic processing skills predicting later symbolic processing skills. Similarly, Wilkey et al. (2022) found that change in nonsymbolic magnitude processing skills from the middle to end of first grade was predicted by symbolic number skills. Goffin and Ansari (2019) suggested that both symbolic and nonsymbolic processing skills may be relevant for arithmetic in young children and may influence each other reciprocally (bidirectional causal influence). The relation between these two constructs is supposed to become refined by the incremental experience with symbols throughout development, so that it tends toward a stronger (or even unidirectional) influence of symbolic skills over nonsymbolic skills in older children. Clearly, these are important theoretical questions for future research. In the age group we investigated, however, the strong association between symbolic and nonsymbolic magnitude comparison tasks can induce redundancy in the regression model, and this redundancy influences the stability of results with respect to these two highly correlated predictors (Field, 2013).

A second issue related to magnitude processing tasks, as pointed out by Caviola et al. (2020), is that the relationship between nonsymbolic magnitude processing measures and mathematical performance might be driven not by causal domain-specific mechanisms but rather by domain-general factors influencing performance in this task. Nonsymbolic magnitude processing performance, for example, has been found to be highly associated with domain-general cognitive dimensions such as fluid intelligence, working memory, and inhibition (Caviola et al., 2020; Gilmore et al., 2013). Accordingly, Wilkey et al. (2022) found that nonsymbolic magnitude processing did not predict unique variance in Grade 1 math achievement once other numerical and non-numerical predictors, including executive function abilities, were included in the model.

Working memory

With respect to the working memory components, we observed a consistent contribution of the central executive to arithmetic in both the concurrent and longitudinal models. The verbal and visuospatial temporary storage components were not significant predictors. These findings are largely in line with recent evidence (Caviola et al., 2020; Clayton et al., 2020; De Vita et al., 2022) and indicate that the working memory component requiring manipulation was the most relevant for attainment in arithmetic. Working memory resources were also strongly related to number writing performance, a finding that replicates previous literature (Camos, 2008; Clayton et al., 2020; Zuber et al., 2009). Given their association with the main predictor of interest (number writing) and the outcome (arithmetic), working memory resources were an important potential confound in our models because they are likely to explain a portion of covariance between arithmetic and number writing that is related to non-numerical memory resources. In the current study, the effect of working memory was accounted for in the regression models. Thus, the robust association between number transcoding and arithmetic we observed can be considered reliable beyond any covariance driven by working memory.

Implications and future directions

Our findings support the notion that transcoding is an important predictor of later arithmetic over and above nonverbal reasoning, working memory, and magnitude comparison skills. Its contributions are comparable across languages with and without number word inversion. We argued that verbal–visual or visual–verbal number conversion as practiced during multi-digit transcoding supports arithmetic competences by training positional knowledge of the Arabic number system. It is important to point out that the observational design of the current study does not allow drawing causal conclusions because the associations we observed may be explained by unmeasured dimensions. Thus, future studies should test the claim that number transcoding is a foundational skill for later arithmetic applying randomized controlled trials with transcoding training. Systematic transcoding training may help children to grasp the place-value system early on, and this might in turn support the acquisition of number processing and arithmetic competences during the first years of primary school.

Data availability

The data of the current study are available at the UK Data Service (https://doi.org/10.5255/UKDASN-854335; https://reshare.ukdataservice.ac.uk/854398/). The SPSS syntax of the main analysis can be retrieved at the Open Science Framework (https://osf.io/p7nzv/).

Acknowledgments

This work was supported by the Economic and Social Research Council (ESRC; ES/N014677/1) and the Austrian Science Fund (FWF; I 2778-G16). The authors applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. We thank Clare Copper, Erin Dysart, Philippa Gibbons, Rebecca Reed, Christina Roberts, Lea Prange, Mariela Rios Diaz, Viktoria Jöbstl and Viktoria Wirth for help with data collection and data entry.

Author contributions

Chiara Banfi: data curation, formal analysis, investigation, project administration, visualization, writing–original draft, and writing–review & editing; Francina J. Clayton: data curation, investigation, project administration, and writing–original draft; Anna F. Steiner: data curation, project administration, investigation, and writing–review & editing; Sabrina Finke: project administration, investigation, and writing–review & editing; Ferenc Kemény: methodology and writing–review & editing; Karin Landerl: conceptualization, funding acquisition, methodology, project administration, resources, supervision, and writing–review & editing; Silke M. Göbel: conceptualization, formal analysis, funding acquisition, methodology, project administration, resources, supervision, writing–original draft, and writing–review & editing.

References

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (pp. 47–89). New York: Academic Press.

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*, 368–394.
Camos, V. (2008). Low working memory capacity impedes both efficiency and learning of number transcoding in children. *Journal of Experimental Child Psychology, 99*, 37–57.
Caviola, S., Colling, L. J., Mammarella, I. C., & Szűcs, D. (2020). Predictors of mathematics in primary school: Magnitude comparison, verbal and spatial working memory measures. *Developmental Science, 23*, e12957.
Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica, 148*, 163–172.
Cheung, P., & Ansari, D. (2021). Cracking the code of place value: The relationship between place and value takes years to master. *Developmental Psychology, 57*, 227–240.
Clayton, F. J., Copper, C., Steiner, A. F., Banfi, C., Finke, S., Landerl, K., & Göbel, S. M. (2020). Two-digit number writing and arithmetic in Year 1 children: Does number word inversion matter?. *Cognitive Development, 56* 100967.
Cragg, L., & Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends in Neuroscience and Education, 3*(2), 63–68.
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.
De Vita, C., Costa, H. M., Tomasetto, C., & Passolunghi, M. C. (2022). The contributions of working memory domains and processes to early mathematical knowledge between preschool and first grade. *Psychological Research Psychologische Forschung, 86*, 497–511.
Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*, 1–42.
Dehaene, S. (2009). Origins of mathematical intuitions: The case of arithmetic. *Annals of the New York Academy of Sciences, 1156*, 232–259.
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..
Dietrich, J. F., Huber, S., Dackermann, T., Moeller, K., & Fischer, U. (2016). Place-value understanding in number line estimation predicts future arithmetic performance. *British Journal of Developmental Psychology, 34*, 502–517.
Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149–1160.
Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology, 123*, 53–72.
Field, A. (2013). *Discovering statistics using IBM SPSS Statistics*. Thousand Oaks, CA: Sage.
Finke, S., Banfi, C., Freudenthaler, H. H., Steiner, A. F., Vogel, S. E., Göbel, S. M., & Landerl, K. (2021). Common and distinct predictors of non-symbolic and symbolic ordinal number processing across the early primary school years. *PLoS ONE, 16*(10), e258847.
Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., ... Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE, 8*(6), e67374.
Göbel, S. M., Watson, S. E., Lervåg, A., & Hulme, C. (2014). Children's arithmetic development: It is number knowledge, not the approximate number sense, that counts. *Psychological Science, 25*, 789–798.
Goffin, C., & Ansari, D. (2019). How are symbols and nonsymbolic numerical magnitudes related? Exploring bidirectional relationships in early numeracy. *Mind, Brain, and Education, 13*, 143–156.
Habermann, S., Donlan, C., Göbel, S. M., & Hulme, C. (2020). The critical role of Arabic numeral knowledge as a longitudinal predictor of arithmetic development. *Journal of Experimental Child Psychology, 193* 104794.
Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology, 103*, 17–29.
Imbo, I., Vanden Bulcke, C., De Brauwer, J., & Fias, W. (2014). Sixty-four or four-and-sixty? The influence of language and working memory on children's number transcoding. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00313.
Lakens, D. (2017). Equivalence tests: A practical primer for *t* tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*, 355–362.
Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science, 1*, 259–269.
Lau, N. T. T., Merkeley, R., Trembley, P., Zhang, S., De Jesus, S., & Ansari, D. (2021). Kindergarteners' symbolic number abilities predict non-symbolic number abilities and math achievement in Grade 1. *Developmental Psychology, 57*, 471–488.
Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in Grades 1–6. *Developmental Science, 17*, 714–726.
Malone, S. A., Burgoyne, K., & Hulme, C. (2020). Number knowledge and the approximate number system are two critical foundations for early arithmetic development. *Journal of Educational Psychology, 112*, 1167–1182.
Meyer, M. L., Salimpoor, V. N., Wu, S. S., Geary, D. C., & Menon, V. (2010). Differential contribution of specific working memory components to mathematics achievement in 2nd and 3rd graders. *Learning and Individual Differences, 20*, 101–109.
Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, H.-C. (2011). Early place-value understanding as a precursor for later arithmetic performance—A longitudinal study on numerical development. *Research in Developmental Disabilities, 32*, 1837–1851.
Moeller, K., Zuber, J., Olsen, N., Nuerk, H. C., & Willmes, K. (2015). Intransparent German number words complicate transcoding—A translingual comparison with Japanese. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00740.
Moura, R., Lopes-Silva, J. B., Vieira, L. R., Paiva, G. M., Prado, A. C. D. A., Wood, G., & Haase, V. G. (2015). From "five" to 5 for 5 minutes: Arabic number transcoding as a short, specific, and sensitive screening tool for mathematics learning difficulties. *Archives of Clinical Neuropsychology, 30*, 88–98.
Moura, R., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., & Haase, V. G. (2013). Transcoding abilities in typical and atypical mathematics achievers: The role of working memory and procedural and lexical competencies. *Journal of Experimental Child Psychology, 116*, 707–727.

Nuerk, H. C., Moeller, K., Klein, E., Willmes, K., & Fischer, M. H. (2011). Extending the mental number line: A review of multi-digit number processing. *Zeitschrift für Psychologie, 219*, 3–22.
Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., ... Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition, 116*, 33–41.
Pickering, S., & Gathercole, S. E. (2001). *Working Memory Test Battery for Children (WMTB-C)*. London: Psychological Corporation.
Pixner, S., Zuber, J., Heřmanová, V., Kaufmann, L., Nuerk, H. C., & Moeller, K. (2011). One language, two number-word systems and many problems: Numerical cognition in the Czech language. *Research in Developmental Disabilities, 32*, 2683–2689.
Poncin, A., Van Rinsveld, A., & Schiltz, C. (2020). Units-first or tens-first: Does language matter when processing visually presented two-digit numbers?. *Quarterly Journal of Experimental Psychology, 73*, 726–738.
Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven's Standard Progressive Matrices and Vocabulary Scales*. Oxford, UK: Oxford Psychologists Press.
Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science, 20*, e12372.
Soltanlou, M., Pixner, S., & Nuerk, H.-C. (2015). Contribution of working memory in multiplication fact network in children may shift from verbal to visuo-spatial: A longitudinal investigation. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01062.
Sowinski, C., LeFevre, J. A., Skwarchuk, S. L., Kamawar, D., Bisanz, J., & Smith-Chant, B. (2015). Refining the quantitative pathway of the pathways to mathematics model. *Journal of Experimental Child Psychology, 131*, 73–93.
Steiner, A. F., Banfi, C., Finke, S., Kemény, F., Clayton, F., Göbel, S. M., & Landerl, K. (2021). Twenty-four or four-and-twenty: Language modulates cross-modal matching for multi-digit numbers in children and adults. *Journal of Experimental Child Psychology, 202* 104970.
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*, 368–387.
Van de Weijer-Bergsma, E., Kroesbergen, E. H., & Van Luit, J. E. H. (2015). Verbal and visual–spatial working memory and mathematical ability in different domains throughout primary school. *Memory & Cognition, 43*, 367–378.
van der Ven, S. H. G., Klaiber, J. D., & van der Maas, H. L. J. (2017). Four and twenty blackbirds: How transcoding ability mediates the relationship between visuospatial working memory and math in a language with inversion. *Educational Psychology, 37*, 487–505.
von Aster, M. G., & Shalev, R. S. (2007). Number development and developmental dyscalculia. *Developmental Medicine and Child Neurology, 49*, 868–873.
Wechsler, D. (2005). *Wechsler individual achievement test – Second UK edition (WIAT-IIUK)*. London, England: Harcourt Assessment.
Wilkey, E. D., Shanley, L., Sabb, F., Ansari, D., Cohen, J. C., Men, V., ... Clarke, B. (2022). Sharpening, focusing, and developing: A study of change in nonsymbolic number comparison skills and math achievement in 1st grade. *Developmental Science, 25*, e13194.
Yuan, L., Prather, R. W., Mix, K. S., & Smith, L. B. (2019). Preschoolers and multi-digit numbers: A path to mathematics through the symbols themselves. *Cognition, 189*, 89–104.
Zuber, J., Pixner, S., Moeller, K., & Nuerk, H.-C. (2009). On the language specificity of basic number processing: Transcoding in a language with inversion and its relation to working memory capacity. *Journal of Experimental Child Psychology, 102*, 60–77.