Journal of Experimental Child Psychology 166 (2018) 232–250

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Developmental trajectories of children's symbolic numerical magnitude processing skills and associated cognitive competencies

Kiran Vanbinst [a,*], Eva Ceulemans [b], Lien Peters [a], Pol Ghesquière [a], Bert De Smedt [a]

[a] Parenting and Special Education Research Unit, Faculty of Psychology and Educational Sciences, University of Leuven, 3000 Leuven, Belgium
[b] Quantitative Psychology and Individual Differences, Faculty of Psychology and Educational Sciences, University of Leuven, 3000 Leuven, Belgium

ARTICLE INFO

Article history:
Received 9 September 2016
Revised 9 July 2017

Keywords:
Symbolic numerical magnitude development
Developmental trajectories
Domain-specific cognitive development
Domain-general cognitive development
Arithmetic development
Longitudinal design

ABSTRACT

Although symbolic numerical magnitude processing skills are key for learning arithmetic, their developmental trajectories remain unknown. Therefore, we delineated during the first 3 years of primary education (5–8 years of age) groups with distinguishable developmental trajectories of symbolic numerical magnitude processing skills using a model-based clustering approach. Three clusters were identified and were labeled as inaccurate, accurate but slow, and accurate and fast. The clusters did not differ in age, sex, socioeconomic status, or IQ. We also tested whether these clusters differed in domain-specific (nonsymbolic magnitude processing and digit identification) and domain-general (visuospatial short-term memory, verbal working memory, and processing speed) cognitive competencies that might contribute to children's ability to (efficiently) process the numerical meaning of Arabic numerical symbols. We observed minor differences between clusters in these cognitive competencies except for verbal working memory for which no differences were observed. Follow-up analyses further revealed that the above-mentioned cognitive competencies did not merely account for the cluster differences in children's development of symbolic numerical magnitude processing skills, suggesting that other factors account for these individual differences. On the other hand, the three trajectories of symbolic numerical magnitude processing revealed remarkable and stable

* Corresponding author.
E-mail address: kiran.vanbinst@kuleuven.be (K. Vanbinst).

https://doi.org/10.1016/j.jecp.2017.08.008
0022-0965/© 2017 Elsevier Inc. All rights reserved.

differences in children's arithmetic fact retrieval, which stresses the importance of symbolic numerical magnitude processing for learning arithmetic.

© 2017 Elsevier Inc. All rights reserved.

Introduction

People are surrounded by Arabic numerical symbols, and numerical literacy has become a crucial skill for everyday life (e.g., Chiswick, Lee, & Miller, 2003; Gerardi, Goette, & Meier, 2013). Numerical literacy is a strong predictor of success at school (Duncan et al., 2007) and of future wealth (Basten, Jaekel, Johnson, Gilmore, & Wolke, 2015). It also has an impact on medical decision making in that it influences, for instance, people's perception of risks and benefits of screenings (Reyna, Nelson, Han, & Dieckmann, 2009). Over the past decade, researchers have investigated numerical literacy by estimating symbolic numerical magnitude processing skills, which reflect the ability to understand the numerical meaning of Arabic digits. Increasing evidence acknowledges the importance of these symbolic numerical magnitude processing skills for learning arithmetic (Dowker, 2005; Gilmore, Attridge, De Smedt, & Inglis, 2014; Jordan, Mulhern, & Wylie, 2009; Price & Fuchs, 2016; Siegler & Lortie-Forgues, 2014; Vanbinst, Ansari, Ghesquière, & De Smedt, 2016; see also De Smedt, Noël, Gilmore, & Ansari, 2013, for a narrative review, and Schneider et al., 2017, for a meta-analysis), but it remains unclear how children develop these symbolic skills.

The dominant view on the development of symbolic magnitude processing skills postulates that these skills are grounded in the ability to represent quantity in a nonsymbolic way (Bugden, DeWind, & Brannon, 2016; Merkley & Ansari, 2016; Siegler & Lortie-Forgues, 2014). Based on the evidence that human infants are able to discriminate between two sets of dots (nonsymbolic representations of quantity) (Xu & Spelke, 2000; Xu, Spelke, & Goddard, 2005), it has been assumed that children progressively learn the numerical meaning of Arabic numerical symbols by connecting these symbolic representations to nonsymbolic representations of quantity. Mundy and Gilmore (2009) specified that the period between 6 and 8 years of age is critical for scaffolding symbolic numerical magnitude representations onto nonsymbolic ones (see also Siegler & Lortie-Forgues, 2014). Studies showing associations between nonsymbolic numerical magnitude processing skills and children's concurrent and future (symbolic) mathematical competence (e.g., Halberda, Mazzocco, & Feigenson, 2008; Libertus, Feigenson, & Halberda, 2011, 2013; Starr, Libertus, & Brannon, 2013) indirectly confirm the connection between nonsymbolic and symbolic numerical magnitude processing. Against this background, nonsymbolic numerical magnitude processing skills are an important determinant of children's early acquisition of symbolic numerical magnitude processing skills (for a review, see Bugden et al., 2016; Merkley & Ansari, 2016; Piazza, 2010; Siegler & Lortie-Forgues, 2014).

Which other *domain-specific cognitive skills* might influence the acquisition of symbolic numerical magnitude processing skills? Clearly, to perform adequately on a symbolic comparison task, it is indispensable that children start with the correct and rapid identification of each presented Arabic numerical symbol (Merkley & Ansari, 2016). Only after identifying both digits can one compare the corresponding numerical quantities and decide on the larger one. Purpura, Baroody, and Lonigan (2013) recently showed that digit identification skills fully mediate the longitudinal association between preschool mathematical abilities of 3- to 5-year-olds and their future mathematical knowledge. The current study aimed to extend these findings by exploring whether individual differences in symbolic numerical magnitude processing might be explained by digit identification skills, which in turn might mediate the association between symbolic numerical magnitude processing and arithmetic. On the other hand, Reeve, Reynolds, Humberstone, and Butterworth (2012) investigated children's numerical development in dot enumeration and symbolic comparison and yet were not able to find associations between this numerical development and speed of identifying Arabic numerical symbols. The differences between these two studies might be explained by differences in age between

the samples under investigation (i.e., kindergarten vs. primary school). Nevertheless, digit identification skills are functionally important for adequately performing on a symbolic comparison task, and for this reason we included digit identification in the current study. Against this background, the current longitudinal study aimed to identify developmental trajectories of symbolic numerical magnitude processing skills during the first 3 years of primary education in order to capture the contribution of children's digit identification skills on this development. By contrasting the distinct trajectories in terms of both nonsymbolic magnitude processing skills and digit identification skills, repeatedly assessed during this developmental period, protracted connections between these domain-specific cognitive competencies and symbolic numerical magnitude processing skills can be established.

Little is known about the extent to which *domain-general cognitive competencies* are needed to develop symbolic numerical magnitude processing skills. Only a few studies to date have investigated interrelations between symbolic numerical magnitude processing and domain-general cognitive competencies. For example, Simmons, Willis, and Adams (2012) observed that visuospatial short-term memory predicted performance on a symbolic magnitude judgment task in which first-graders needed to choose the largest of three symbolic numbers. These researchers explained this association by assuming that children employ visuospatial numerical representations when comparing (multidigit) numbers. In the same age group, Xenidou-Dervou, van Lieshout, and van der Schoot (2013) found that the performance on a symbolic approximate addition task was significantly related to visuospatial short-term memory (as well as verbal working memory). Similar results were obtained in third- and fourth-graders by Caviola, Mammarella, Cornoldi, and Lucangeli (2012). By contrast, Träff (2013) could not detect significant associations between fifth- and sixth-graders' performance on a symbolic comparison task and their visuospatial short-term memory or verbal working memory. There is, to the best of our knowledge, no research available that investigated the contribution of these domain-general competencies to the *learning* of symbolic magnitude processing over a longer developmental time period. This was precisely one of the goals of the current longitudinal study.

Existing research using symbolic comparison tasks to capture children's symbolic numerical magnitude processing skills typically relied on children's speed of comparing Arabic numerical symbols for investigating associations between symbolic comparison and arithmetic (see De Smedt et al., 2013, for a review). When drawing conclusions on associations with speeded measures, it is crucial to control for general effects of speed because this appears to correlate with academic achievement (Koponen, Salmi, Eklund, & Aro, 2013). The current study, therefore, included a measure of children's rapid automatized naming (RAN) skills. RAN is known to involve children's processing speed as well as their capacity to retrieve (phonological) information from long-term memory (van den Bos, Zijlstra, & Spelber, 2002; van den Bos, Zijlstra, & Van den Broeck, 2003). It can be predicted that RAN contributes to children's symbolic numerical magnitude processing skills because it helps children to efficiently retrieve numerical magnitude information from long-term memory. The potential association between RAN and symbolic numerical magnitude processing skills was further explored in this longitudinal study.

To understand individual differences in the development of children's symbolic numerical magnitude processing, we applied a model-based clustering approach on longitudinal data. Surprisingly, only a small number of studies in the field of mathematical cognition have used such a model-based clustering approach. For example, this approach has been used to reveal subtypes of dyscalculia (Bartelet, Ansari, Vaessen, & Blomert, 2014; von Aster, 2000) and to capture profiles of individual differences in children's arithmetic fact development (Vanbinst, Ceulemans, Ghesquière, & De Smedt, 2015). This latter study in third- to fifth-graders identified profiles of arithmetic fact development that persistently differed in symbolic numerical magnitude processing skills over time regardless of differences in age, sex, socioeconomic status, nonverbal reasoning, general mathematics achievement, and reading ability. These longitudinal data highlight the long-lasting importance of symbolic numerical magnitude processing skills even beyond the early grades of primary school. Reeve et al. (2012) conducted a study in which they performed consecutive cluster analyses on children's symbolic comparison speed assessed at all grades of primary education. The same three clusters were detected at each time point throughout primary education, and the majority of participants remained in the same cluster across the study, indicating that individual differences in symbolic numerical magnitude processing skills (as well as dot enumeration) remain relatively stable over time.

In contrast to Reeve et al. (2012), in which a cluster analysis was conducted at successive time points, the current longitudinal study applied a cluster analysis over the course of time in order to capture distinguishable trajectories of individual differences in development. We focused on the first 3 years of primary education because we expected the development of symbolic magnitude processing to be the largest in this age group. At the start of each grade, we administered a symbolic magnitude comparison task, and the accuracy and speed at each time point were used as parameters for the model-based cluster analysis. To further characterize the clusters, we compared them on the abovementioned domain-specific (i.e., nonsymbolic numerical magnitude processing) skills and digit identification skills and domain-general (i.e., visuospatial short-term memory, verbal working memory, and RAN) cognitive competencies that might contribute to individual differences in children's symbolic numerical magnitude processing development.

To give a complete view of these developmental trajectories of symbolic numerical magnitude processing skills, we compared the clusters on their arithmetic development (i.e. their acquisition and reliance on arithmetic fact retrieval). There exist strong associations between symbolic numerical magnitude processing and arithmetic fact retrieval during the initial stages of development (Bartelet, Vaessen, Blomert, & Ansari, 2014; Vanbinst, Ghesquière, & De Smedt, 2015) as well as the later stages of development (Vanbinst et al., 2015) (see Schneider et al., 2017, for a meta-analysis), and recent evidence even highlighted that symbolic numerical magnitude processing is as important to arithmetical learning as phonological awareness is to the acquisition of reading (Vanbinst et al., 2016). Against this background (see also Dowker, 2005; Gilmore et al., 2014; Jordan et al., 2009; Price & Fuchs, 2016), we also compared the identified developmental trajectories of symbolic magnitude processing on arithmetic in second and third grades, expecting these trajectories to differ in terms of their arithmetic competence.

Method

Participants

Participants were recruited from an ongoing longitudinal study in which three Belgian schools participated. The initial sample comprised 88 first-graders at Time Point 1. At Time Point 2 a year later, symbolic comparison data were available for 67 participants of the initial sample. At the start of third grade, Time Point 3, symbolic comparison data were available for 51 participants. Because we wanted to estimate children's development in symbolic numerical magnitude processing from first grade to third grade, we decided to include only those participants whose symbolic comparison data were available at each time point. These data were available for only 51 children. Missing data were due to illness at one of the time points but also to changing schools and to repeating or skipping a year of primary school. We also performed sensitivity analyses to determine whether the current findings were not unduly biased by this attrition. Missing completely at random (MCAR) analyses on symbolic comparison data (accuracy and response time) missing at the second and third time points revealed that data were missing at random (Little's MCAR test: chi-square = 1.32, $df = 2$, $p = .936$). MCAR analyses on arithmetic data (accuracy, response time, and frequency fact retrieval) missing at the second and third time points also showed that data were missing at random (Little's MCAR test: chi-square = 2.125, $df = 3$, $p = .547$). We subsequently compared the data of children in the final sample ($n = 51$) and those with missing data ($n = 37$) on the cognitive measures assessed in first grade (Time Point 1). These analyses revealed that there were no significant group differences on these measures between children with missing data and those without missing data.

All participants ($n = 51$) of the final sample (29 girls and 22 boys; $M_{age} = 6$ years 2 months, $SD = 4$ months at Time Point 1) were native Dutch speakers and came from middle- to upper middle-class families. None of them repeated a grade, and written informed parental consent was obtained for all of them.

Materials

Materials were computer tasks designed with E-Prime 1.0 software (Schneider, Eschmann, & Zuccolotto, 2002), paper-and-pencil tasks, and standardized tests.

Symbolic numerical magnitude processing

Children's symbolic numerical magnitude processing skills were measured with a classic comparison task. In this task, children needed to compare two simultaneously presented Arabic digits displayed on either side of a 15-inch computer screen. They needed to indicate the larger of those two Arabic digits by pressing a key on the side of the larger one. Stimuli comprised all combinations of digits 1–9, yielding 72 trials. The position of the largest digit was counterbalanced. Each trial was initiated by the experimenter and started with a central 200-ms fixation point, followed by a blank of 800 ms. Stimuli appeared 1000 ms after trial initiation and remained visible until response. Response times and answers were registered. To familiarize children with the key assignments, three practice trials were presented. Children's mean response time and accuracy were used in the analyses as an indication of their task performance. Split-half reliability of this task calculated in the current sample was .87 for accuracy and .94 for response time.

Domain-specific cognitive competencies

Nonsymbolic numerical magnitude processing

Children's nonsymbolic numerical magnitude processing skills were measured with a comparison task analogous to the one used to assess symbolic numerical magnitude processing. The Arabic digits, used in the symbolic test format, were replaced by dot arrays. The nonsymbolic stimuli were generated with the MATLAB script provided by Piazza, Izard, Pinel, Le Bihan, and Dehaene (2004) and were controlled for non-numerical parameters such as dot size, total occupied area, and density. On one half of the trials dot size, array size, and density were positively correlated with number, and on the other half of the trials dot size, array size, and density were negatively correlated. This was done to avoid decisions being dependent on non-numerical cues or visual features. A trial started with a 200-ms fixation point in the center of the screen. Stimuli appeared after 1000 ms and disappeared again after 840 ms to avoid counting the number of dots. Each trial was initiated by the experimenter with a control key. Response times and answers were registered by the computer. To familiarize children with the key assignments, three practice trials were included per task. For the analyses, we used children's mean response time and accuracy on the task as an indication of their performance. Split-half reliability of this task calculated in the current sample was .57 for accuracy and .78 for response time.

Digit identification

In the digit identification task, each of the numbers 1 to 9 was successively presented twice on the computer screen. Children were asked to name each digit. Response time was registered by a voice key, after which the answer was entered on the keyboard by the experimenter. There were two practice trials to make children familiar with task administration. In the subsequent analyses, we used children's mean response time and accuracy to indicate their task performance. Split-half reliability of this task calculated in the current sample was .72 for accuracy and .84 for response time.

Domain-general cognitive competencies

Verbal working memory

A classic listening span task from the Working Memory Test Battery for Children (Pickering & Gathercole, 2001), adapted to Dutch (see De Smedt et al., 2009, for more elaborated task details), was used to administer verbal working memory. In this non-numerical task, children needed to judge the correctness of a series of recorded sentences (true vs. false). They were also instructed to memorize the last word in every sentence and to recall those words in the presented order at the end of each trial. The task started at a list length of one, and three trials of each list length were presented. If children recalled at least two of three trials of the same list length correctly, the list length was increased by one sentence. The total score on this task equaled the number of trials recalled correctly. Reliability of this task in a sample of Flemish children of the same age was .64 (De Smedt et al., 2009).

Visuospatial short-term memory

The Corsi block task from the Working Memory Test Battery for Children (Pickering & Gathercole, 2001) was used to assess children's visuospatial short-term memory (see De Smedt et al., 2009, for more elaborated task details). For each trial, the experimenter tapped out a sequence, at a rate of one block per second, on a board with nine blocks. Children were instructed to reproduce the sequence in the correct order. The task started with trials at a list length of two blocks, and three trials of each list length were presented. The list length was increased by one block if children recalled at least two of three trials of the same list length correctly. If children failed to do this, the task was terminated. A trial was scored as correct if all stimuli of that trial were recalled in the correct order. The task yielded a total score equal to the number of trials recalled correctly. Reliability of this task in a sample of Flemish children of the same age was .77 (De Smedt et al., 2009).

Rapid automatized naming

RAN was evaluated with a rapid automatic color naming task (van den Bos et al., 2003). The presented test card consisted of 50 stimuli (5 columns of 10 stimuli), with each color (black, blue, red, yellow, and green) appearing 10 times. Prior to testing, children were required to name the stimuli in the last column to determine whether they were familiar with all of the presented stimuli. The total time to name all stimuli on the card was recorded for each task and used as an indicator of children's task performance. Reliability of this task derived from the manual was .88 (van den Bos, 2004).

Arithmetic

Arithmetic and strategy use were assessed by means of an addition and subtraction task. Stimuli were selected from the so-called standard set of single-digit arithmetic problems (LeFevre, Sadesky, & Bisanz, 1996), which excludes tie problems (e.g., 6 + 6) and problems containing 0 or 1 as operand or answer. Only one of each pair of commutative problems was selected, resulting in a set of 28 problems per operation. The position of the largest operand was counterbalanced. Children were asked to perform both accurately and quickly. Responses were verbal. A voice key registered children's response time, after which the experimenter recorded the answer. Children could use whatever strategy they wanted to use. On a trial-by-trial basis, the experimenter asked children to verbally report the strategy they used to solve the arithmetic problem. Similar to other studies in arithmetic (e.g., Imbo & Vandierendonck, 2007; Torbeyns, Verschaffel, & Ghesquière, 2004), strategies were classified as *retrieval* (i.e., if children expressed that they immediately knew the answer and if, at the same time, there was no evidence of overt calculations), *procedural* (i.e., if children indicated that they used counting or decomposed the problem into smaller sub-problems to arrive at the solution), or *other* (i.e., if children did not know how they solved the problem). This classification method is a valid and reliable way of assessing children's arithmetic strategy use (Siegler & Stern, 1998). Two practice trials were presented to familiarize children with task administration. Split-half reliability of the addition task calculated in this sample was .63 for accuracy and .86 for response time, and split-half reliability of the subtraction task was .70 for accuracy and .96 for response time.

Nonverbal reasoning

Nonverbal reasoning was included as a control measure and was assessed with Raven's Standard Progressive Matrices (Raven, Court, & Raven, 1992). For each child, a standardized score ($M = 100$, $SD = 15$) was calculated. Reliability for this test in Flemish children of the same age is .90 (De Smedt et al., 2009).

Procedure

All tasks were individually administered in a quiet room at participants' own school except for Raven's matrices, which was group based. Task order was fixed for all participants. At the start of first grade (Time 1: September 2011), second grade (Time 2: September 2012), and third grade (Time 3: September 2013), all participants completed the symbolic comparison task, the nonsymbolic

comparison task, and the digit identification task. The listening span task, the Corsi block task, and the color naming task were assessed only at primary school entrance (Time 1). The arithmetic task was administered at Times 2 and 3. Nonverbal reasoning was assessed in December of first grade.

Results

The first section of Results covers the cluster analysis and descriptive statistics of each identified cluster. The subsequent section discusses cluster differences in the above-mentioned domain-specific and domain-general cognitive competencies that were predicted to support the development of children's symbolic numerical magnitude processing skills. The last section focuses on differences between clusters in their arithmetic development and reliance on fact retrieval. At relevant places, we reported partial eta-squared values as measure of effect size. Partial eta-squared values range between 0 and 1 and can be interpreted as follows: .02 ∼ small, .13 ∼ medium, and .26 ∼ large (e.g., Pierce, Block, & Aguinis, 2004).

Symbolic comparison trials for which children had a response time lower than 300 ms or higher than 5000 ms were discarded (<3% of all trials at Time 1 and <2% of all trials at Times 2 and 3). The same data trimming was applied for nonsymbolic comparison (<3% of all trials at Times 1, 2, and 3) and digit identification (<2% of all trials at Times 1, 2, and 3) tasks. For arithmetic, trials deviating more than 3 standard deviations from each individual's mean response time and trials with a response time below 500 ms (<2% of all trials at Times 1 and 2) were excluded. Trials with incorrect responses or incorrect voice-key registrations (<5% of all trials) were also excluded from response time analyses. To account for multiple comparisons, all reported results were analyzed with Bonferroni corrections.

Cluster analysis

Children's symbolic numerical magnitude processing skills were estimated by considering their performance patterns on a symbolic comparison task throughout the first 3 years of primary school. This performance on a symbolic comparison task was evaluated by the mean accuracy and mean response time at the start of each grade. We applied a model-based cluster analysis (Banfield & Raftery, 1993) on these two parameters estimated at three different time points by using the MCLUST package of Fraley and Raftery (2007), implemented in R.[1] Model-based clustering (Banfield & Raftery, 1993) models the data as a mixture of, in this case, multivariate normal distributions. The clusters correspond to the distributions, which differ in terms of means and/or covariance structure. Different variants are available that impose more or less restrictions on the covariance structure (e.g., variables are uncorrelated, variances are identical, covariances are equal across clusters). The means and variance–covariance matrices of the clusters are estimated during the clustering as well as the prior probabilities of belonging to a cluster. Based on these estimates and the observed data of a person, the posterior probabilities are computed, indicating for each cluster how likely it is that this person was drawn from that cluster. Each person is assigned to the cluster for which the posterior probability is highest (Bayes' rule). Model-based clustering is often preferred over heuristic (i.e. non-model-based) clustering methods, such as *K*-means clustering and hierarchical cluster analysis, because the principled statistical approach allows one to use well-known inferential tools. For example, to decide how many clusters should be used and which restrictions can be imposed on the covariance matrices, one can apply the popular Bayesian information criterion (BIC; Schwarz, 1978; see also Bulteel, Wilderjans, Tuerlinckx, & Ceulemans, 2013), which balances the log likelihood of the data given the model, number of parameters (prior probabilities, means, and covariances, taking imposed set of restrictions into account), and sample size. Concretely, a comparison of the BIC values of models with 1 to 9 clusters revealed that the model that divided our group of participants into three clusters had the lowest BIC value (i.e., −1662.475) and, hence, provided

[1] We should note that we did not apply growth curve modeling in which time would be included as a predictor, with differences in intercept and slope determining cluster membership. Our aim was not to estimate individual growth curves but rather to investigate whether children could be clustered in terms of their performance patterns (i.e., response time and accuracy at each time point, not their change) on the symbolic comparison task across the 3 years under study. In other words, we measured two variables at each time point—response time and accuracy—and wanted to study their interrelations over time.

the best fit to the data. The selected model is of the VEE type, implying that the multivariate normal distributions that underlie the three clusters differ in volume but have the same shape and orientation. These three identified clusters were considered as three distinguishable developmental trajectories of symbolic numerical magnitude processing skills and were labeled as *inaccurate* ($n = 10$), *accurate but slow* ($n = 25$), and *accurate and fast* ($n = 16$). Fig. 1 displays for each identified cluster the mean accuracy and response time on the symbolic comparison task per grade.

We calculated a 3 × 3 repeated-measures analysis of variance (ANOVA) with grade (1 vs. 2 vs. 3) as a within-participant factor and cluster (inaccurate vs. accurate but slow vs. accurate and fast) as a between-participants factor on the mean accuracy and response time of the symbolic comparison task. With regard to accuracy, there was a main effect of grade, $F(2, 96) = 10.92$, $p < .001$, $\eta_p^2 = .185$; the three clusters became more accurate over time. There was a main effect of cluster, $F(2, 48) = 16.14$, $p < .001$, $\eta_p^2 = .402$; the inaccurate cluster was significantly less accurate than the accurate but slow cluster ($p < .001$) and the accurate and fast cluster ($p < .001$). The latter two did not differ in terms of accuracy ($p = .851$). There was no Grade × Cluster interaction, $F(4, 96) = 2.08$, $p = .090$.

The analysis of the response time revealed a main effect of grade, $F(2, 96) = 224.32$, $p < .001$, $\eta_p^2 = .824$, demonstrating that all children from all three clusters became faster with progressing time (all ps $< .001$). There was a main effect of cluster, $F(2, 48) = 20.45$, $p < .001$, $\eta_p^2 = .460$; the accurate and fast cluster was systematically faster than the accurate but slow cluster ($p < .001$) and the inaccurate cluster ($p < .001$). These last two clusters did not differ ($p = .177$). Grade interacted with cluster membership, $F(4, 96) = 5.59$, $p < .001$, $\eta_p^2 = .189$; there was a striking difference at the start of each grade between the accurate and fast and accurate but slow clusters. Differences between the accurate and fast and inaccurate clusters changed over time; significant differences between these two emerged in first and third grades but not in second grade. Children from the accurate and fast cluster accelerated with progressing time, but children from the inaccurate cluster became remarkably faster between first and second grades yet made only limited progress in speed between second and third grades.

Table 1 presents the detailed descriptive statistics of the three identified clusters. This table illustrates that the clusters did not differ in terms of age, $F(2, 50) = 1.39$, $p = .259$, sex, $\chi^2(2) = 1.68$, $p = .432$, and nonverbal reasoning, $F(2, 47) = 2.06$, $p = .140$. All subsequent analyses were repeated with nonverbal reasoning as a covariate, but this did not change our findings. All children's mothers were asked to report their educational level as a marker of socioeconomic status (2 missing values), and no differences in socioeconomic status were observed between clusters, $\chi^2(2) = 5.23$, $p = .073$.

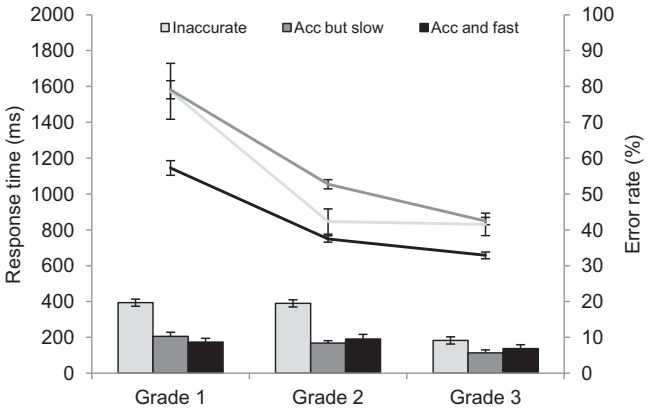

Fig. 1. Developmental trajectories of symbolic numerical magnitude processing skills per cluster across grades. Lines represent response time on the left *y* axis, and bars depict error rates on the right *y* axis. Error bars represent 1 standard deviation of the mean. Acc, accurate.

Table 1
Descriptive statistics of the identified clusters.

	Inaccurate	Accurate but slow	Accurate and fast
Age[a]	6.23 (0.28)	6.15 (0.32)	6.31 (0.32)
Sex	4 boys, 6 girls	9 boys, 16 girls	9 boys, 7 girls
Nonverbal reasoning[b]	97.90 (12.41)	109.13 (14.65)	105.27 (16.38)
Mother's educational level[c]	7/3	7/17	5/10

Note. Standard deviations are in parentheses.
[a] Mean age in years at primary school entrance.
[b] IQ score on Raven's Standard Progressive Matrices.
[c] Primary or secondary/higher education.

Cluster differences on domain-specific cognitive competencies

Nonsymbolic numerical magnitude processing

The mean accuracy and response time on the nonsymbolic comparison task are displayed in Fig. 2 for each cluster per grade. We calculated a 3 × 3 repeated-measures ANOVA with grade (1 vs. 2 vs. 3) as a within-participant factor and cluster (inaccurate vs. accurate but slow vs. accurate and fast) as a between-participants factor on accuracy and response time. The analysis of accuracy showed a significant main effect of grade, $F(2, 96) = 19.03$, $p < .001$, $\eta_p^2 = .284$. Post hoc t tests demonstrated that children's accuracy increased significantly from first grade to second grade ($p < .001$) but not from second grade to third grade ($p = 1.00$). We observed no main effect of cluster, $F(2, 48) = 2.88$, $p = .065$, and no Grade × Cluster interaction, $F(4, 96) = 1.22$, $p = .307$. With regard to response time, there was a main effect of grade, $F(2, 96) = 21.19$, $p < .001$, $\eta_p^2 = .306$, showing that all clusters became faster over time, as well as a main effect of cluster, $F(2, 48) = 3.57$, $p = .036$, $\eta_p^2 = .129$. Post hoc t tests revealed no differences between the inaccurate cluster and both the accurate and fast ($p = .102$) and accurate but slow ($p = 1.00$) clusters. There was a marginally significant difference between the accurate and fast cluster and the accurate but slow cluster ($p = .06$). Grade interacted with cluster, $F(4, 96) = 2.63$, $p = .039$, $\eta_p^2 = .099$. At primary school entrance, there was no significant difference between the accurate and fast cluster and the accurate but slow cluster ($p = .100$). Differences between both clusters surprisingly increased over time; the accurate and fast cluster appeared to be significantly faster than the accurate but slow cluster at the start of second ($p = .034$) and third ($p = .020$) grades.

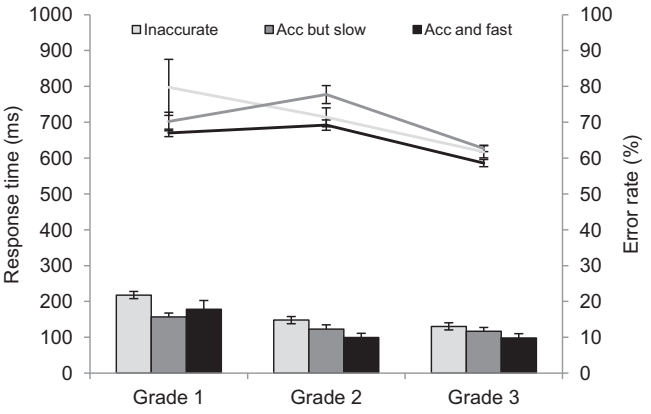

Fig. 2. Mean response time and error rate (% errors) per cluster on the nonsymbolic comparison task across grades. Lines represent response time on the left y axis, and bars depict error rates on the right y axis. Error bars represent 1 standard deviation of the mean. Acc, accurate.

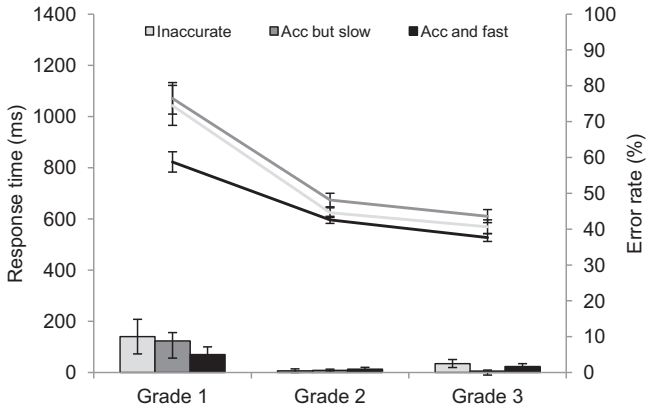

Fig. 3. Mean response time and error rate (% errors) per cluster on the digit identification task across grades. Lines represent response time on the left *y* axis, and bars depict error rates on the right *y* axis. Error bars represent 1 standard error of the mean. Acc, accurate.

Digit identification

Fig. 3 displays the mean accuracy and response time on the digit identification task per cluster across grades. A 3 × 3 repeated-measures ANOVA with grade (1 vs. 2 vs. 3) as a within-participant factor and cluster (inaccurate vs. accurate but slow vs. accurate and fast) as a between-participants factor was conducted on accuracy and response time. With regard to accuracy, there was a main effect of grade, $F(2, 94) = 14.91$, $p < .001$, $\eta_p^2 = .241$. Post hoc *t* tests demonstrated that children's accuracy for identifying digits increased significantly from first grade to second grade ($p < .001$) but not from second grade to third grade ($p = .185$) due to ceiling effects. There was no main effect of cluster, $F(2, 47) = 0.490$, $p = .616$, and no Grade × Cluster interaction, $F(4, 94) = 0.743$, $p = .565$. The analysis of the response time revealed a main effect of grade, $F(2, 94) = 122.95$, $p < .001$, $\eta_p^2 = .723$; all three clusters became faster over time (all $ps < .005$). Post hoc *t* tests, unpacking the main effect of cluster, $F(2, 47) = 4.64$, $p = .014$, $\eta_p^2 = .165$, showed that only the accurate but slow and accurate and fast clusters differed significantly; the accurate but slow cluster was always slower ($p = .012$). The Grade × Cluster interaction was not significant, $F(4, 94) = 2.42$, $p = .054$, although the contrast between the accurate but slow and accurate and fast clusters was significant in first grade ($p = .013$) but diminished over time (Grade 2: $p = .065$; Grade 3: $p = .052$).

Cluster differences on domain-general cognitive competencies

The descriptive statistics of children's verbal working memory, visuospatial short-term memory, and RAN are displayed in Table 2. This table also presents an overview of cluster differences on these domain-general cognitive competencies.

Verbal working memory

Table 2 indicates that no cluster differences emerged on the listening span task.

Visuospatial short-term memory

Significant cluster differences emerged on the Corsi block task (see Table 2). Post hoc *t* tests revealed that the *i*naccurate cluster performed significantly lower on the Corsi block task than the accurate and fast cluster, $t(24) = -2.969$, $p = .007$, $d = -1.24$. No differences were observed between the inaccurate and accurate but slow clusters, $t(33) = -1.712$, $p = .096$, or between the accurate but slow and accurate and fast clusters, $t(39) = 1.817$, $p = .077$.

Table 2
Descriptive statistics of verbal working memory, visuospatial short-term memory, and RAN skills per cluster.

Task	Inaccurate	Accurate but slow	Accurate and fast	F	p	η_p^2	Post hoc cluster differences
	M (SD)	M (SD)	M (SD)				
Verbal working memory							
Listening span	3.40 (1.71)	3.52 (1.61)	3.19 (1.11)	0.242	.786	.010	I = AS = AF
Visuospatial short-term memory							
Corsi block	5.80 (2.70)	7.20 (1.96)	8.25 (1.53)	4.621	.015	.161	I < AF I = AS AS = AF
RAN							
Color naming	62.50 (10.17)	74.92 (20.23)	59.25 (12.64)	4.955	.011	.171	AS < AF I = AS I = AF

Note. I, inaccurate; AS, accurate but slow; AF, accurate and fast.

Rapid automatized naming

The three clusters differed in RAN at primary school entrance (see Table 2). Post hoc t tests indicated that the accurate but slow cluster performed significantly slower ($p = .014$) on a color naming task than the accurate and fast cluster. No further cluster differences were observed.

Do differences in cognitive competencies explain differences in symbolic development?

To verify whether cluster differences in children's development of symbolic numerical magnitude processing skills were not merely accounted for by the above-mentioned cognitive competencies, we recalculated the 3 × 3 repeated-measures ANOVA with grade (1 vs. 2 vs. 3) as a within-participant factor and cluster (inaccurate vs. accurate but slow vs. accurate and fast) as a between-participants factor on the mean accuracy and response time of the symbolic comparison task while also controlling for these cognitive competencies. The results of these analyses are presented in Table 3. Concretely, the cognitive competencies for which significant cluster differences were observed (i.e., nonsymbolic magnitude processing, digit identification, visuospatial short-term memory, and RAN) were systematically entered as covariates in order to check the impact of these cognitive competencies on the separate developmental trajectories of symbolic numerical magnitude processing skills.

Table 3 illustrates that after also controlling for the effects of cognitive competencies, cluster differences in symbolic comparison accuracy and response time remained significant. Turning to the effects of the covariates, the majority of partial eta-squared values for these covariates represented small effects (RAN, visuospatial short-term memory, nonsymbolic comparison accuracy, and digit identification accuracy) and indicated that these variables did not have an important role in explaining cluster differences in symbolic magnitude processing. The covariates nonsymbolic comparison response time and digit identification response time showed medium and strong effects, respectively. This suggests that these two variables played a role in children's symbolic magnitude processing skills, yet they did not fully explain the observed cluster differences.

Arithmetic (fact) development

Performance on the addition and subtraction task was strongly correlated (all rs > .71), and to improve clarity the reported results were averaged across operations. Children's arithmetic development was estimated by considering three parameters of arithmetic skills (i.e., mean accuracy, mean response time, and fact retrieval frequency) at the start of second and third grades (see Table 4).

A 2 × 3 repeated-measures ANOVA with grade (2 vs. 3) as a within-participant factor and cluster (inaccurate vs. accurate but slow vs. accurate and fast) as a between-participants factor was conducted on accuracy, response time, and frequencies of fact retrieval. With regard to accuracy, there

Table 3
Overview of changes in cluster effects on the mean accuracy and response time of the symbolic comparison task after including covariates (i.e., nonsymbolic magnitude processing, digit identification, visuospatial short-term memory, and RAN).

	F	p	η_p^2
Symbolic comparison accuracy			
Main cluster effect without covariate	16.41	<.001	.402
Main cluster effect with covariate nonsymbolic comparison accuracy[a]	12.12	<.001	.340
Covariate effect: Nonsymbolic comparison accuracy[a]	4.059	.050	.079
Main cluster effect with covariate digit identification accuracy[a]	14.85	<.001	.392
Covariate effect: Digit identification accuracy[a]	2.551	.117	.053
Main cluster effect with covariate visuospatial short-term memory	12.41	<.001	.346
Covariate effect: Visuospatial short-term memory	1.857	.179	.038
Main cluster effect with covariate RAN	16.81	<.001	.417
Covariate effect: RAN	1.216	.276	.025
Symbolic comparison response time			
Main cluster effect without covariate	20.45	<.001	.460
Main cluster effect with covariate nonsymbolic response time[b]	16.22	<.001	.408
Covariate effect: Nonsymbolic comparison response time[b]	24.512	<.001	.343
Main cluster effect with covariate digit identification response time[b]	12.52	<.001	.353
Covariate effect: Digit identification response time[b]	5.974	.018	.115
Main cluster effect with covariate visuospatial short-term memory	17.54	<.001	.427
Covariate effect: Visuospatial short-term memory	1.736	.194	.036
Main cluster effect with covariate RAN	14.66	<.001	.384
Covariate effect: RAN	4.390	.042	.085

[a] The included covariate consisted of the mean accuracy averaged across all time points.
[b] The included covariate consisted of the mean response time averaged across all time points.

Table 4
Acquisition of arithmetic skills per cluster.

	Inaccurate	Accurate but slow	Accurate and fast
Mean accuracy (% correct)			
Grade 2	81.79 (14.15)	88.13 (10.77)	90.10 (6.48)
Grade 3	88.25 (10.13)	92.31 (4.46)	93.11 (4.75)
Mean response time (ms)			
Grade 2	6137.35 (2056.71)	6106.29 (2510.11)	4248.77 (1246.36)
Grade 3	5054.45 (1685.82)	4325.48 (1457.73)	3084.45 (979.08)
Fact retrieval frequency			
Grade 2	15.00 (7.77)	22.28 (3.82)	23.44 (6.16)
Grade 3	23.60 (6.74)	22.00 (6.79)	27.69 (7.09)

Note. Standard deviations are in parentheses.

was a main effect of grade, $F(1, 48) = 8.61$, $p = .005$, $\eta_p^2 = .152$, indicating that all clusters improved in terms of accuracy over developmental time. There was no significant main effect of cluster, $F(2, 48) = 3.103$, $p = .054$, $\eta_p^2 = .114$. No Grade × Cluster interaction was observed, $F(2, 48) = 0.344$, $p = .711$, $\eta_p^2 = .014$.

The analysis of the response time revealed a main effect of grade, $F(1, 48) = 21.76$, $p < .001$, $\eta_p^2 = .312$, demonstrating that the three clusters became faster at solving arithmetic over time (all ps < .001). There was a main effect of cluster, $F(1, 48) = 6.96$, $p = .002$, $\eta_p^2 = .225$; the accurate and fast cluster was systematically faster than the inaccurate ($p = .007$) and accurate but slow ($p = .007$) clusters. The inaccurate and accurate but slow clusters did not differ ($p = .100$). This effect of cluster remained after taking into account RAN, $F(2, 47) = 5.45$, $p = .007$, $\eta_p^2 = .188$. Grade did not interact with cluster, $F(2, 48) = 0.732$, $p = .486$.

Concerning the repeated-measures ANOVA on the mean frequencies of fact retrieval, this analysis revealed a main effect of grade, $F(1, 48) = 9.70$, $p = .003$, $\eta_p^2 = .168$, suggesting that all clusters increasingly relied on fact retrieval with progressing time. There was a main effect of cluster, $F(2, 48) = 6.77$,

$p = .003$, $\eta_p^2 = .220$; post hoc t tests revealed that the accurate and fast cluster tended to retrieve more arithmetic facts from memory than the inaccurate ($p = .002$) and accurate but slow ($p = .052$) clusters. In turn, the inaccurate and accurate but slow clusters did not differ ($p = .259$). Grade interacted with cluster, $F(2, 48) = 3.77$, $p = .03$, $\eta_p^2 = .136$; cluster differences became smaller over time but remained significant at the start of second grade, $F(2, 50) = 8.09$, $p = .001$, $\eta_p^2 = .252$, as well as third grade, $F(2, 50) = 3.37$, $p = .043$, $\eta_p^2 = .123$.

Discussion

Numerous studies have highlighted the importance of symbolic numerical magnitude processing skills for learning arithmetic (Dowker, 2005; Jordan et al., 2009; Siegler & Lortie-Forgues, 2014; see also Schneider et al., 2017, for a meta-analysis), and it has even been suggested that symbolic numerical magnitude processing is a good candidate for screening children at risk for developing mathematical difficulties (Brankaer, Ghesquière, & De Smedt, 2017; Nosworthy, Bugden, Archibald, Evans, & Ansari, 2013). Nevertheless, it remains largely unclear how children develop these symbolic numerical magnitude processing skills. The current longitudinal study identified three clusters or groups of children who differed in their development of symbolic numerical magnitude processing skills throughout the first 3 years of primary education. These groups with different developmental trajectories of symbolic numerical magnitude processing skills did not differ in terms of age, sex, socioeconomic status, and nonverbal reasoning and were labeled as *inaccurate, accurate but slow,* and *accurate and fast.* We also tested whether these groups differed in domain-specific and domain-general cognitive competencies that might contribute to children's ability to (efficiently) process the numerical meaning of Arabic numerical symbols. The results showed minor differences between clusters in nonsymbolic numerical magnitude processing skills, digit identification skills, visuospatial short-term memory, and RAN but showed no cluster differences in verbal working memory. Additional analyses revealed that cluster differences in children's development of symbolic numerical magnitude processing skills were not merely accounted for by the above-mentioned domain-specific and domain-general cognitive competencies, indicating that further studies are needed to determine the origins of individual differences in symbolic numerical magnitude processing. On the other hand, the three trajectories of symbolic numerical magnitude processing revealed remarkable and stable differences in children's arithmetic fact retrieval, replicating the stable association between symbolic numerical magnitude processing and arithmetic (Dowker, 2005; Gilmore et al., 2014; Jordan et al., 2009; Price & Fuchs, 2016).

The three identified developmental trajectories were marked by differences in children's development of symbolic numerical magnitude processing skills from the onset of primary education up to third grade. With progressing time, the symbolic numerical magnitude processing skills of all participants became more accurate, but compared with children with an accurate but slow or accurate and fast developmental trajectory, children with an inaccurate developmental trajectory always made more errors when solving the symbolic comparison task. All three clusters improved in terms of speed when solving the symbolic comparison task, which was not explained by differences in speed, considered in terms of children's RAN. Compared with children with an accurate but slow or inaccurate trajectory, children with an accurate and fast trajectory solved the symbolic comparison particularly fast. In summary, these results highlight that striking individual differences exist in children's ability to access symbolic numerical magnitude information from long-term memory and that these individual differences persist over the first 3 years of primary education.

After ascertaining the existence of three trajectories in children's symbolic numerical magnitude processing development, we contrasted the clusters on both domain-specific and domain-general cognitive competencies that (possibly) help to learn and process the numerical meaning of Arabic numerical symbols. In view of the idea that symbolic representations are grounded in preexisting nonsymbolic representations of quantity (Bugden et al., 2016; Merkley & Ansari, 2016; Siegler & Lortie-Forgues, 2014), we predicted that clusters would also differ in terms of their performance on a nonsymbolic comparison task. Surprisingly, our longitudinal data did not reveal striking cluster differences in nonsymbolic numerical magnitude processing. This finding, along with a few studies

(Lyons, Nuerk, & Ansari, 2015; Matejko & Ansari, 2016; Sasanguie, De Smedt, Defever, & Reynvoet, 2012; Vanbinst, Ghesquière, & De Smedt, 2012) showing weak or nonsignificant correlations between performance on symbolic and nonsymbolic comparison tasks, questions the prominent role of nonsymbolic numerical magnitude processing as a ground for developing symbolic numerical magnitude processing skills. This aligns with other observations suggesting that children learn to connect Arabic numerical symbols with their corresponding magnitudes independently from earlier acquired nonsymbolic numerical magnitude processing skills.

First, in children with dyscalculia—a specific learning disorder in mathematics—deficits in symbolic numerical magnitude processing have consistently and persistently been observed, but deficits in nonsymbolic numerical magnitude processing have, by contrast, been found in only some, but not all, studies on dyscalculia (e.g., Mazzocco, Feigenson, & Halberda, 2011; Mussolin, Mejias, & Noël, 2010; Piazza et al., 2010), particularly not in younger children (e.g., De Smedt & Gilmore, 2011; Iuculano, Tang, Hall, & Butterworth, 2008; Landerl & Kölle, 2009; Rousselle & Noël, 2007). If nonsymbolic abilities are foundational for the acquisition of symbolic numerical magnitude processing skills, then nonsymbolic impairments in dyscalculia should be the most prominent at the youngest age groups, which is not the case. Second, strong associations between nonsymbolic numerical magnitude processing and mathematics achievement have not been found systematically (see Schneider et al., 2017, for a meta-analysis). Finally, longitudinal evidence by Sasanguie, Defever, Maertens, and Reynvoet (2014) illustrated that nonsymbolic numerical magnitude processing skills in preschool did not predict future symbolic numerical magnitude processing skills. Even more surprising, Mussolin, Nys, Content, and Leybaert (2014) found that nonsymbolic numerical magnitude processing skills of 3- and 4-year-olds were predicted by their previously measured cardinality proficiency as well as their symbolic number knowledge. The reverse prediction was, however, not significant.

These findings, together with those observed in the current study, contradict the assumption that symbolic representations are grounded in preexisting nonsymbolic representations and even suggest a reversed association, namely that children's early symbolic numerical magnitude processing skills might optimize their nonsymbolic numerical magnitude processing skills. Indeed, the current data show that differences between our clusters in nonsymbolic comparison are not present at primary school entrance but only start to emerge in second and third grades, when children from the accurate and fast cluster performed significantly faster compared with children from the accurate but slow cluster. It might be that once children have acquired a certain level of skill in symbolic numerical magnitude processing, they start to use their symbolic knowledge to solve the nonsymbolic comparison task. It is important to mention that this study began at only the earliest years of primary education, and hence the start of formal mathematics instruction, a period when children learn to fluently process the numerical meaning of Arabic numerical symbols. It remains possible that, before formal math instruction, such as when 2- to 4-year-old preschoolers are acquainted with Arabic numerical symbols for the very first time, children initially connect these symbolic representations with the corresponding nonsymbolic representation of quantity, and that this connection fades quickly away with time. This idea matches with a recent study by Matejko and Ansari (2016) in which the authors explored whether developmental trajectories of symbolic and nonsymbolic numerical magnitude processing skills relate to each other during the first year of formal schooling. Distinct symbolic versus nonsymbolic trajectories were captured, but both trajectories were related only during the first 6 months of primary education. The study by Matejko and Ansari, as well as the current study, points out that the link between nonsymbolic and symbolic numerical magnitude processing might not be unidirectional as was originally thought. Developmental research on bidirectional relations between symbolic and nonsymbolic numerical magnitude processing skills across time is needed to further unravel this issue.

Digit identification skills of all three clusters were contrasted, revealing that children with an accurate but slow trajectory identified Arabic digits systematically slower than children with an accurate but fast trajectory. It is not surprising that this observation was particularly prominent in first grade but less so during the ensuing years. Covariance analyses further illustrated that digit identification skills did not account for the identification of three clusters with distinct developmental trajectories of symbolic numerical magnitude processing skills. These findings allow us to conclude that

protracted individual differences in children's symbolic numerical magnitude processing development cannot be reduced to individual differences in digit identification skills.

The role of working memory in learning arithmetic, as well as in other learning processes such as reading, has received a lot of research attention over the past decades (see Peng, Namkung, Barnes, & Sun, 2015, for a meta-analysis). By contrast, very few studies have examined whether and how working memory is related to symbolic numerical magnitude processing (Caviola et al., 2012; Simmons et al., 2012; Xenidou-Dervou et al., 2013). This is surprising because it is not unlikely that working memory resources are necessary for learning to process the numerical meaning of symbolic representations such as Arabic numerical symbols. In the current study, we could not find cluster differences in verbal working memory, but significant differences between developmental trajectories did occur on visuospatial short-term memory. More specifically, children with an inaccurate trajectory performed significantly lower on a Corsi block task than children with an accurate and fast trajectory. Additional covariance analyses showed, however, that the capacity of children's visuospatial short-term memory did not fully explain the differences between the identified clusters. Prior studies on the association between visuospatial short-term memory and symbolic numerical magnitude processing were conducted in the same age group (Caviola et al., 2012; Simmons et al., 2012; Xenidou-Dervou et al., 2013). This link may be explained by the age of the participants under study, indicating that visuospatial short-term memory is important at the earlier stages of children's symbolic numerical magnitude processing development. This is also in line with recent research with children with Turner syndrome, a genetic syndrome characterized by impairments in visuospatial short-term memory and dyscalculia (Brankaer, Ghesquière, De Wel, Swillen, & De Smedt, 2016). This study revealed that in this genetic condition visuospatial short-term memory was strongly related to impairments in symbolic numerical magnitude processing skills, although these associations were much less strong or even absent in controls or children with other genetic disorders and dyscalculia such as 22q11 deletion syndrome. Taken together, this suggests that visuospatial short-term memory can be important in symbolic numerical magnitude processing and its impairments, although further studies are needed to clarify for which children this conclusion holds. It might be that comparison of (symbolic) magnitudes requires the employment of visuospatial representations (in memory) because numerical representations of magnitude may be visuospatial in nature as they are represented on a mental number line (Simmons et al., 2012). This, then, might explain why individual differences in visuospatial short-term memory correlate with individual differences in symbolic magnitude comparison (Simmons et al., 2012). On the other hand, alternative explanations, which are related to the specific nature of the Corsi block task, are possible. Specifically, it could be that children use some counting strategy on the Corsi block task in order to maintain the blocks in their memory because the tapping of the blocks might look like the counting of objects. This somewhat numerical nature of the Corsi block task might explain the observed association between symbolic numerical magnitude processing and visuospatial short-term memory. Such an explanation remains speculative, however, and future studies are needed to further test these possibilities. Moreover, the current data do not speak to the relevance of other visuospatial skills, and this clearly also represents an area for future study.

Similar to Reeve et al. (2012), the current study demonstrated that participant variability in the development of symbolic numerical magnitude processing skills is not reducible to general effects of processing speed. We found that children with an accurate but slow trajectory of symbolic numerical magnitude processing development performed slower on a RAN task when entering primary school compared with children with an accurate and fast trajectory. RAN skills reflect not only children's general processing speed but also their ability to retrieve information from long-term memory such as numerical magnitude information needed to solve a comparison task. It is not surprising that children from the accurate but slow trajectory experience more difficulties in retrieving this information from long-term memory. Additional analyses showed that long-lasting individual differences in children's symbolic numerical magnitude processing development remained after accounting for RAN skills, indicating that the influence of RAN skills on the development of symbolic magnitude processing is only limited.

This longitudinal study supports prior studies, signaling a link between proficient symbolic numerical magnitude processing skills and successful arithmetic (fact) development (e.g., Baroody, 2006; Bartelet et al., 2014; Jordan, Hanich, & Kaplan, 2003; Vanbinst et al., 2015). All children enhanced their

mastery of arithmetic facts from second grade to third grade, but especially in the accurate and fast trajectory did children solve additions and subtractions more frequently with fact retrieval, and they were also systematically faster than children in the two other trajectories irrespective of cluster differences in RAN. Not surprisingly, children from the inaccurate cluster formed the group of children who remained slow and continued making errors while solving arithmetic. It was also striking that this group used very few facts to solve arithmetic, especially at the start of second grade. Our longitudinal evidence of early primary education clearly suggests that symbolic numerical magnitude processing plays an important role in children's development of arithmetic fact retrieval (for similar results, see Bartelet et al., 2014; Geary, Hoard, & Bailey, 2012; see also Baroody, 2006; Jordan et al., 2003). Numerous studies also point to behavioral associations between children's RAN and their growth in mathematics, more specifically their acquisition of arithmetic facts (Koponen, Georgiou, Salmi, Leskinen, & Aro, 2016). In the current study, cluster differences in arithmetic development remained after RAN skills were taken into account, suggesting that RAN seems to play a less prominent role in children's arithmetic fact development compared with simultaneously considered (symbolic) numerical magnitude processing skills.

When evaluating the current findings, it is important to note that they were based on a rather small sample size due to the fact that we included only those participants for which we had data over the entire period of 3 years. Therefore, it is important to replicate this study with a larger sample size. On the other hand, we post hoc calculated the power of our analyses given the current sample size ($n = 51$) and assuming three groups of participants, an alpha level of .05, and a medium effect size, and these calculations indicated that our study had power of 0.77. We also reanalyzed the data with Bayesian statistics—for which there is no longer a need for power analysis because the probability of different hypotheses (including the null) is evaluated (see Dienes, 2011, p. 276, for an excellent elaboration of this issue)—and these additional analyses converged to the same results as with the originally reported frequentist analyses. It is also important to highlight that the current sample is restricted to children from middle- to upper middle-class neighborhoods. Future studies, therefore, should focus on samples that are more varied in terms of social backgrounds.

In conclusion, cluster differences in symbolic numerical magnitude processing development remain despite taking into account important domain-specific and domain-general cognitive competencies. Collectively, these findings stress the importance of symbolic numerical magnitude processing skills for learning arithmetic. On the other hand, they also raise the question of whether this association might be bidirectional. Future studies should investigate this by examining children's symbolic numerical processing development even before children receive formal mathematics instruction, in order to explore how this symbolic development predicts proficiency in different components of arithmetic (Cowan et al., 2011; Dowker, 2005; Jordan et al., 2009). Future research should also further investigate the origins of individual differences in symbolic numerical magnitude processing. Potential candidates of other domain-specific cognitive competencies include spontaneous focusing on numerosity (Hannula, Lepola, & Lehtinen, 2010) and knowledge of counting procedures and/or counting principles (e.g., Fazio, Bailey, Thompson, & Siegler, 2014; Goffin & Ansari, 2016). In addition, even noncognitive determinants, such as home numeracy, should be considered in future research because there is evidence to suggest that home experiences before schooling are important in understanding the development of symbolic numerical magnitude processing skills (LeFevre et al., 2009). Nevertheless, the current study highlights the existence of long-lasting individual differences in symbolic numerical magnitude processing that coexist with remarkable and stable individual differences in learning arithmetic.

References

Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics, 49*, 803–821.

Baroody, A. (2006). Why children have difficulties mastering the basic number combinations and how to help them. *Teaching Children Mathematics, 13*(1), 22–31.

Bartelet, D., Ansari, D., Vaessen, A., & Blomert, L. (2014). Cognitive subtypes of mathematics learning difficulties in primary education. *Research in Developmental Disabilities, 35*, 657–670.

Bartelet, D., Vaessen, A., Blomert, L., & Ansari, D. (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *Journal of Experimental Child Psychology, 117*, 12–28.

Basten, M., Jaekel, J., Johnson, S., Gilmore, C., & Wolke, D. (2015). Preterm birth and adult wealth: Mathematics skills count. *Psychological Science, 26*, 1608–1619.
Brankaer, C., Ghesquière, P., & De Smedt, B. (2017). Symbolic magnitude processing in elementary school children: A group administered paper-and-pencil measure (SYMP Test). *Behavior Research Methods, 49*, 1361–1373.
Brankaer, C., Ghesquière, P., De Wel, A., Swillen, A., & De Smedt, B. (2016). Numerical magnitude processing impairments in genetic syndromes: A cross-syndrome comparison of Turner and 22q11.2 deletion syndromes. *Developmental Science*. https://doi.org/10.1111/desc.12458. Advance online publication.
Bugden, S., DeWind, N. K., & Brannon, E. M. (2016). Using cognitive training studies to unravel the mechanisms by which the approximate number system supports symbolic math ability. *Current Opinion in Behavioral Sciences, 10*, 73–80.
Bulteel, K., Wilderjans, T. F., Tuerlinckx, F., & Ceulemans, E. (2013). CHull as an alternative to AIC and BIC in the context of mixtures of factor analyzers. *Behavior Research Methods, 45*, 782–791.
Caviola, S., Mammarella, I. C., Cornoldi, C., & Lucangeli, D. (2012). The involvement of working memory in children's exact and approximate mental addition. *Journal of Experimental Child Psychology, 112*, 141–160.
Chiswick, B. R., Lee, Y. L., & Miller, P. W. (2003). Schooling, literacy, numeracy, and labor market success. *Economic Record, 79*, 165–181.
Cowan, R., Donlan, C., Shepherd, D. L., Cole-Fletcher, R., Saxton, M., & Hurry, J. (2011). Basic calculation proficiency and mathematics achievement in elementary school children. *Journal of Educational Psychology, 103*, 786–803.
De Smedt, B., & Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology, 108*, 278–292.
De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology, 103*, 186–201.
De Smedt, B., Noël, M., Gilmore, C., & Ansari, D. (2013). The relationship between symbolic and non-symbolic numerical magnitude processing and the typical and atypical development of mathematics: Evidence from brain and behavior. *Trends in Neuroscience and Education, 2*, 48–55.
Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives in Psychological Science, 6*, 274–290.
Dowker, A. (2005). *Individual differences in arithmetic: Implications for psychology, neuroscience, and education*. Hove, UK: Psychology Press.
Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Sexton, H. (2007). School readiness and later achievement. *Developmental psychology, 43*(6), 1428.
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.
Fraley, C., & Raftery, A. E. (2007). Model-based methods of classification: Using the MCLUST software in chemometrics. *Journal of Statistical Software, 18*, paper i06.
Geary, D. C., Hoard, M. K., & Bailey, D. H. (2012). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *Journal of Learning Disabilities, 45*, 291–307.
Gerardi, K., Goette, L., & Meier, S. (2013). Numerical ability predicts mortgage default. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 11267–11271.
Gilmore, C., Attridge, N., De Smedt, B., & Inglis, M. (2014). Measuring the approximate number system in children: Exploring the relationships among different tasks. *Learning and Individual Differences, 29*, 50–58.
Goffin, C., & Ansari, D. (2016). Beyond magnitude: Judging ordinality of symbolic number is unrelated to magnitude comparison and independently relates to individual differences in arithmetic. *Cognition, 150*, 68–76.
Halberda, J., Mazzocco, M. M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*, 665–668.
Hannula, M. M., Lepola, J., & Lehtinen, E. (2010). Spontaneous focusing on numerosity as a domain-specific predictor of arithmetical skills. *Journal of Experimental Child Psychology, 107*, 394–406.
Imbo, I., & Vandierendonck, A. (2007). The role of phonological and executive working memory resources in simple arithmetic strategies. *European Journal of Cognitive Psychology, 19*, 910–933.
Iuculano, T., Tang, J., Hall, C. W. B., & Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Developmental Science, 11*, 669–680.
Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology, 85*, 103–119.
Jordan, J. A., Mulhern, G., & Wylie, J. (2009). Individual differences in trajectories of arithmetical development in typically achieving 5- to 7-year-olds. *Journal of Experimental Child Psychology, 103*, 469–479.
Koponen, T., Georgiou, G., Salmi, P., Leskinen, M., & Aro, M. (2016). A meta-analysis of the relation between RAN and mathematics. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000182. Advance online publication.
Koponen, T., Salmi, P., Eklund, K., & Aro, T. (2013). Counting and RAN: Predictors of arithmetic calculation and reading fluency. *Journal of Educational Psychology, 105*, 162–175.
Landerl, K., & Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. *Journal of Experimental Child Psychology, 103*, 546–565.
LeFevre, J., Sadesky, G. S., & Bisanz, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology, 22*, 216–230.
LeFevre, J., Skwarchuk, S., Smith-Chant, B. L., Fast, L., Kamawar, D., & Bisanz, J. (2009). Home numeracy experiences and children's math performance in the early school years. *Canadian Journal of Behavioural Science, 41*, 55–66.
Libertus, M. E., Feigenson, L., & Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Developmental Science, 14*, 1292–1300.
Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *Journal of Experimental Child Psychology, 116*, 829–838.

Lyons, I. M., Nuerk, H.-C., & Ansari, D. (2015). Rethinking the implications of numerical ratio effects for understanding the development of representational precision and numerical processing across formats. *Journal of Experimental Psychology: General, 144*, 1021–1035.
Matejko, A. A., & Ansari, D. (2016). Trajectories of symbolic and nonsymbolic magnitude processing in the first year of formal schooling. *PLoS ONE, 11*(3), e149863.
Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82*, 1224–1237.
Merkley, R., & Ansari, D. (2016). Why numerical symbols count in the development of mathematical skills: Evidence from brain and behavior. *Current Opinion in Behavioral Sciences, 10*, 14–20.
Mundy, E., & Gilmore, C. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology, 103*, 490–502.
Mussolin, C., Mejias, S., & Noël, M. (2010). Symbolic and non-symbolic number comparison in children with and without dyscalculia. *Cognition, 115*, 10–25.
Mussolin, C., Nys, J., Content, A., & Leybaert, J. (2014). Symbolic number abilities predict later approximate number system acuity in preschool children. *PLoS ONE, 9*(3), e91839.
Nosworthy, N., Bugden, S., Archibald, L., Evans, B., & Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLoS ONE, 8*(7), e67918.
Peng, P., Namkung, J., Barnes, M., & Sun, C. (2015). 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*, 455–473.
Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences, 14*, 542–551.
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.
Piazza, M., Izard, V., Pinel, P., Le Bihan, D., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron, 44*, 547–555.
Pickering, S., & Gathercole, S. (2001). *Working memory test battery for children (WMTB-C)*. London: Psychological Corporation.
Pierce, C. A., Block, C. A., & Aguinis, H. (2004). Cautionary note on reporting eta-squared values from multifactor ANOVA designs. *Educational and Psychological Measurement, 64*, 916–924.
Price, G. R., & Fuchs, L. S. (2016). The mediating relation between symbolic and nonsymbolic foundations of math competence. *PLoS ONE, 11*(2), e148981.
Purpura, D. J., Baroody, A. J., & Lonigan, C. J. (2013). The transition from informal to formal mathematical knowledge; Mediation by numeral knowledge. *Journal of Educational Psychology, 105*, 453–464.
Raven, J. C., Court, J. H., & Raven, J. (1992). *Standard progressive matrices*. Oxford, UK: Oxford Psychologists Press.
Reeve, R., Reynolds, F., Humberstone, J., & Butterworth, B. (2012). Stability and change in markers of core numerical competencies. *Journal of Experimental Psychology: General, 141*, 649–666.
Reyna, V. F., Nelson, W. L., Han, P. K., & Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychological Bulletin, 135*, 943–973.
Rousselle, L., & Noël, M. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs. non-symbolic number magnitude. *Cognition, 102*, 361–395.
Sasanguie, D., De Smedt, B., Defever, E., & Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *British Journal of Developmental Psychology, 30*, 344–357.
Sasanguie, D., Defever, E., Maertens, B., & Reynvoet, B. (2014). The approximate number system is not predictive for symbolic number processing in kindergartners. *Quarterly Journal of Experimental Psychology, 67*, 271–280.
Schneider, M., Beeres, K., Coban, L., Merz, S., 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*. https://doi.org/10.1111/desc.12372. Advance online publication.
Schneider, W., Eschmann, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.
Siegler, R. S., & Lortie-Forgues, H. (2014). An integrative theory of numerical development. *Child Development Perspectives, 8*, 144–150.
Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology, 127*, 377–397.
Simmons, F. R., Willis, C., & Adams, A. (2012). Different components of working memory have different relationships with different mathematical skills. *Journal of Experimental Child Psychology, 111*, 139–155.
Starr, A., Libertus, M. E., & Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 18116–18120.
Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2004). Strategic aspects of simple addition and subtraction: The influence of mathematical ability. *Learning and Instruction, 14*, 177–195.
Träff, U. (2013). The contribution of general cognitive abilities and number abilities to different aspects of mathematics in children. *Journal of experimental child psychology, 116*(2), 139–156.
van den Bos, K. (2004). *Serieel benoemen en woorden lezen* [Serial naming and word reading]. Groningen, The Netherlands: University of Groningen.
van den Bos, K. P., Zijlstra, B. J. H., & Spelber, H. C. L. (2002). Life-span data on continuous-naming speeds of numbers, letters, colors and pictured objects, and word-reading speed. *Scientific Studies of Reading, 6*, 25–49.
van den Bos, K. P., Zijlstra, B. J. H., & Van den Broeck, W. (2003). Specific relations between alphanumeric-naming speed and reading speeds of monosyllabic and multisyllabic words. *Applied Psycholinguistics, 24*, 407–430.
Vanbinst, K., Ansari, D., Ghesquière, P., & De Smedt, B. (2016). Symbolic magnitude processing is as important to arithmetic development as phonological awareness is to reading. *PLoS ONE, 11*(3), e151045.

Vanbinst, K., Ceulemans, E., Ghesquière, P., & De Smedt, B. (2015). Profiles of children's arithmetic fact development: A model-based clustering approach. *Journal of Experimental Child Psychology, 133*, 29–46.
Vanbinst, K., Ghesquière, P., & De Smedt, B. (2012). Numerical magnitude representations and individual differences in children's arithmetic strategy use. *Mind, Brain, and Education, 6*, 129–136.
Vanbinst, K., Ghesquière, P., & De Smedt, B. (2015). Does numerical processing uniquely predict first graders' future development of single-digit arithmetic? *Learning and Individual Differences, 37*, 153–160.
von Aster, M. (2000). Developmental cognitive neuropsychology of number processing and calculation: Varieties of developmental dyscalculia. *European Child & Adolescent Psychiatry, 9*, 41–57.
Xenidou-Dervou, I., van Lieshout, E. C. D. M., & van der Schoot, M. (2013). Working memory in nonsymbolic approximate arithmetic processing: A dual-task study with preschoolers. *Cognitive Science, 38*, 101–127.
Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*, B1–B11.
Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science, 8*, 88–101.