frontiers
RESEARCH TOPICS

NUMERICAL DEVELOPMENT - FROM COGNITIVE FUNCTIONS TO NEURAL UNDERPINNINGS

Topic Editors
Korbinian Moeller, Elise Klein,
Klaus Willmes and Karin Kucian

frontiers in PSYCHOLOGY

ABOUT FRONTIERS

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

FRONTIERS JOURNAL SERIES

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing.

All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

DEDICATION TO QUALITY

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

WHAT ARE FRONTIERS RESEARCH TOPICS?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area!

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

NUMERICAL DEVELOPMENT - FROM COGNITIVE FUNCTIONS TO NEURAL UNDERPINNINGS

Topic Editors:
Korbinian Moeller, Knowledge Media Research Center, Germany
Elise Klein, Knowledge Media Research Center, Germany
Klaus Willmes, University Hospital Aachen, RWTH Aachen University, Germany
Karin Kucian, University Childrens Hospital Zurich, Switzerland

Living at the beginning of the 21st century requires being numerate, because numerical abilities are not only essential for life prospects of individuals but also for economic interests of post-industrial knowledge societies. Thus, numerical development is at the core of both individual as well as societal interests. There is the notion that we are already born with a very basic ability to deal with small numerosities. Yet, this often called "number sense" seems to be very restricted, approximate, and driven by perceptual constraints. During our numerical development in formal (e.g., school) but also informal contexts (e.g., family, street) we acquire culturally developed abstract symbol systems to represent exact numerosities – in particular number words and Arabic digits – refining our numerical capabilities.

In recent years, numerical development has gained increasing research interest documented in a growing number of behavioural, neuro-scientific, educational, cross-cultural, and neuropsychological studies addressing this issue. Additionally, our understanding of how numerical competencies develop has also benefitted considerably from the advent of different neuro-imaging techniques allowing for an evaluation of developmental changes in the human brain. In sum, we are now starting to put together a more and more coherent picture of how numerical competencies develop and how this development is associated with neural changes as well. In the end, this knowledge might also lead to a better understanding of the reasons for atypical numerical development which often has grieve consequences for those who suffer from developmental dyscalculia or mathematics learning disabilities.

Therefore, this Research Topic deals with all aspects of numerical development: findings from behavioural performance to underlying neural substrates, from cross-sectional to longitudinal evaluations, from healthy to clinical populations. To this end, we included empirical contributions using different experimental methodologies, but also theoretical contributions, review articles, or opinion papers.

Table of Contents

Numerical development—from cognitive functions to neural underpinnings

Korbinian Moeller[1,2]*, Elise Klein[1,3], Karin Kucian[4] and Klaus Willmes[3]

[1] Neurocognition Lab, Knowledge Media Research Center, Tuebingen, Germany
[2] Department of Psychology, University of Tuebingen, Germany
[3] Section Neuropsychology, Department of Neurology, University Hospital Aachen, RWTH Aachen University, Aachen, Germany
[4] Center for MR-Research, University Children's Hospital Zurich, Zurich, Switzerland
*Correspondence: korbinian.moeller@uni-tuebingen.de

Living at the beginning of the 21st century requires being numerate because numerical abilities are not only essential for life prospects of individuals but also for economic interests of post-industrial knowledge societies (e.g., Butterworth et al., 2011). In recent years, numerical development has gained increasing research interest. Following this trend, we invited empirical and theoretical contributions for a Research Topic on *Numerical development—from cognitive functions to neural underpinnings*. We are grateful to all authors for their high-quality contributions, the reviewers for their constructive comments and suggestions in the interactive peer-review, and the publisher's editorial team for their excellent support.

The different contributions nicely illustrate that the construct *numerical development* does not denote a unitary, clearly circumscribed, and comprehensive entity. Instead, the empirical, review, opinion, and commentary articles clearly suggest that it is important to consider different empirical and theoretical perspectives evaluating cross-domain (e.g., language or spatial abilities) but also domain-specific [e.g., basic numerical competencies, approximate number system (ANS), spatial-numerical associations (SNA)] determinants of and influences on typical but also atypical numerical development.

A first set of studies investigated cross-domain as well as domain-specific influences on *typical numerical development*. With respect to cross-domain influences LeFevre et al. (2013) showed a reliable impact of children's spatial abilities on numerical skills. Additionally, Durkin et al. (2013) observed that language ability is a unique predictor of actual and future numerical achievement. Getting closer to the domain of numerical cognition, Lafay et al. (2013) found that finger counting may be useful but not necessary to develop accurate symbolic numerical competencies. As regards basic numerical precursor competencies, two studies investigated the influence of the ANS on numerical development. Mejias and Schiltz (2013) suggest that the ANS may be targeted by educational strategies as it seems to be associated with socio-economic status. Lonnemann et al. (2013) observed that children's addition performance is associated with different markers of the ANS during development. For secondary school children, Huber et al. (2013) found the relation of multiplication and division to be stronger for easier problems and more skilled

(i.e., higher grade) students. Moreover, two fMRI studies investigated neural correlates of numerical development. Mussolin et al. (2013) observed differential developments of the contributions of the right and left intraparietal sulcus (IPS) to magnitude comparison over age, while Gullick and Wolford (2013) found a fronto-parietal shift of activation with age in general, but also specific effects on the lateralization of IPS involvement for processing negative numbers. Finally, Lambrechts et al. (2013) observed that processing the continuous quantities time and space seems resilient to healthy aging similar to numerosity.

With a particular focus on the determinants of *atypical numerical processing* in mathematics learning disability (MLD) or developmental dyscalculia (DD), Mazzocco et al. (2013) showed that specific basic whole number misconceptions reliably predict atypical performance on Grade 8 arithmetic tests. Furthermore, Chu et al. (2013) found that even though inacuity of the ANS is a reliable predictor of risk for MLD it may not be its primary source. Nevertheless, Landerl (2013) evaluated the influences of other basic numerical competencies on numerical development and observed that children with DD exhibited specific impairments. Importantly, however, Haase et al. (2014) concluded that subtypes of MLD may not only be associated with content-related deficits but also more general impairments of information processing should be considered. In line with this view, Van Viersen et al. (2013) suggested that evaluating eye-fixation behavior may provide interesting new insights into the mechanisms underlying DD. In their Opinion, Kaufmann et al. (2013) argue that MLD/DD is a heterogeneous disorder resulting from individual differences in development or function at neuroanatomical, neuropsychological, behavioral, and interactional levels. Finally, Käser et al. (2013) described the development and first successful evaluation of a multimodal and adaptive computer-based training program for children with DD.

Another set of studies investigated interactions in the *processing of numbers and space*. With respect to SNA, Knops et al. (2013) observed a reversed operational momentum (OM) effect in children with their degree of attentional control predicting the propensity to exhibit the OM effect. Kucian et al. (2013) commented that the left–right associations underlying the OM effect may dependent on development, reflecting an interaction of

visuo-spatial and attentional processes with number related skills which might account for the non-observation of an OM effect in children (with DD). Moreover, Link et al. (2014) observed that unbounded number line estimation may be a valuable tool for assessing primary school children's spatial representation of number magnitude in an unbiased manner. Goldman et al. (2013) argued that the development of an analog comparison process and the specific processing of end stimuli contribute to the emergence of the mental number line. In their review, Patro et al. (2014) suggested a taxonomy for the classification of SNA from infancy to late preschool years. On the other hand, another two studies addressed direct interactions between the processing of numerical and physical magnitude as reflected by the size congruity effect (Henik and Tzelgov, 1982). Leibovich et al. (2013) found that numerical and physical magnitudes are represented by different, yet interactive systems. Ben-Shalom et al. (2013) observed that even preschoolers were able to process number magnitude information automatically. Finally, unrelated to space, Gabriel et al. (2013) suggested that conceptual knowledge about *fractions* and procedural knowledge about how to manipulate them should be distinguished. Faulkenberry (2013) commented that the evaluation of solution strategies might be beneficial to differentiate procedural and conceptual knowledge.

Extending the focus on *language influences* Imbo et al. (2014) and Klein et al. (2013) specifically investigated influences of number word inversion on numerical development. They found that language, but not working memory capacity, predicted the number of inversion errors and conclude that inversion-related difficulties do not fade over time, respectively. Lopes-Silva et al. (2014) observed the more basic perceptual phonemic awareness to predict number transcoding reliably, whereas magnitude processing and working memory did not.

As documented by this broad range of studies dealing with different aspects of numerical development—from behavioral performance to underlying neural substrates, from cross-sectional to longitudinal evaluations, from healthy to clinical populations—the current Research Topic brought together the expertise of researchers from different backgrounds and clearly advanced our understanding of numerical development—a topic with both scientific and every-day relevance.

ACKNOWLEDGMENTS

REFERENCES

Ben-Shalom, T., Berger, A., and Henik, A. (2013). My brain knows numbers! – an ERP study of preschoolers' numerical knowledge. *Front. Psychol.* 4:716. doi: 10.3389/fpsyg.2013.00716

Butterworth, B., Varma, S., and Laurillard, D. (2011). Dyscalculia: from brain to education. *Science* 332, 1049–1053. doi: 10.1126/science.1201536

Chu, F. W., van Marle, K., and Geary, D. C. (2013). Quantitative deficits of preschool children at risk or mathematical learning disability. *Front. Psychol.* 4:195. doi: 10.3389/fpsyg.2013.00195

Durkin, K., Mok, P. L. H., and Conti-Ramsden, G. (2013). Severity of specific language impairment predicts delayed development in number skills. *Front. Psychol.* 4:581. doi: 10.3389/fpsyg.2013.00581

Faulkenberry, T. J. (2013). The conceptual/procedural distinction belongs to strategies, not tasks: a comment on Gabriel et al. (2013). *Front. Psychol.* 4:820. doi: 10.3389/fpsyg.2013.00820

Gabriel, F., Coché, F., Szucs, D., Carette, V., Rey, B., and Content, A. (2013). A componential view of children's difficulties in learning fractions. *Front. Psychol.* 4:715. doi: 10.3389/fpsyg.2013.00715

Goldman, R., Tzelgov, J., Ben-Shalom, T., and Berger, A. (2013). Two separate processes affect the development of the mental number line. *Front. Psychol.* 4:317. doi: 10.3389/fpsyg.2013.00317

Gullick, M. M., and Wolford, G. (2013). Understanding less than nothing: children's neural response to negative numbers shifts across age and accuracy. *Front. Psychol.* 4:584. doi: 10.3389/fpsyg.2013.00584

Haase, V. G., Júlio-Costa, A., Lopes-Silva, J. B., Starling-Alves, I., Antunes, A. M., Pinheiro-Chagas, P., et al. (2014). Contributions from specific and general factors to unique deficits: two cases of mathematics learning difficulties. *Front. Psychol.* 5:102. doi: 10.3389/fpsyg.2014.00102

Henik, A., and Tzelgov, J. (1982). Is three greater than five: the relation between physical and semantic size in comparison tasks. *Mem. Cognit.* 10, 389–395.

Huber, S., Fischer, U., Moeller, K., and Nuerk, H.-C. (2013). On the interrelation of multiplication and division in secondary school children. *Front. Psychol.* 4:740. doi: 10.3389/fpsyg.2013.00740

Imbo, I., Vanden Bulcke, C., De Brauwer, J., and Fias, W. (2014). Sixty-four or four-and-sixty? The influence of language and working memory on children's number transcoding. *Front. Psychol.* 5:313. doi: 10.3389/fpsyg.2014.00313

Käser, T., Baschera, G.-M., Kohn, J., Kucian, K., Richtmann, V., Grond, U., et al. (2013). Design and evaluation of the computer-based training program Calcularis for enhancing numerical cognition. *Front. Psychol.* 4:489. doi: 10.3389/fpsyg.2013.00489

Kaufmann, L., Mazzocco, M. M., Dowker, A., von Aster, M., Göbel, S. M., Grabner, R. H., et al. (2013). Dyscalculia from a developmental and differential perspective. *Front. Psychol.* 4:516. doi: 10.3389/fpsyg.2013.00516

Klein, E., Bahnmueller, J., Mann, A., Pixner, S., Kaufmann, L., Nuerk, H.-C., et al. (2013). Language influences on numerical development - inversion effects on multi-digit number processing. *Front. Psychol.* 4:480. doi: 10.3389/fpsyg.2013.00480

Knops, A., Zitzmann, S., and McCrink, K. (2013). Examining the presence and determinants of operational momentum in childhood. *Front. Psychol.* 4:325. doi: 10.3389/fpsyg.2013.00325

Kucian, K., Plangger, F., O'Gorman, R., and von Aster, M. (2013). Operational momentum effect in children with and without developmental dyscalculia. *Front. Psychol.* 4:847. doi: 10.3389/fpsyg.2013.00847

Lafay, A., Thevenot, C., Castel, C., and Fayol, M. (2013). The role of fingers in number processing in young children. *Front. Psychol.* 4:488. doi: 10.3389/fpsyg.2013.00488

Lambrechts, A., Karolis, V., Garcia, S., Obende, J., and Cappelletti, M. (2013). Age does not count: resilience of quantity processing in healthy ageing. *Front. Psychol.* 4:865. doi: 10.3389/fpsyg.2013.00865

Landerl, K. (2013). Development of numerical processing in children with typical and dyscalculic arithmetic skills - a longitudinal study. *Front. Psychol.* 4:459. doi: 10.3389/fpsyg.2013.00459

LeFevre, J.-A., Jimenez Lira, C., Sowinski, C., Cankaya, O., Kamawar, D., and Skwarchuk, S.-L. (2013). Charting the role of the number line in mathematical development. *Front. Psychol.* 4:641. doi: 10.3389/fpsyg.2013.00641

Leibovich, T., Diesendruck, L., Rubinsten, O., and Henik, A. (2013). The importance of being relevant: modulation of magnitude representations. *Front. Psychol.* 4:369. doi: 10.3389/fpsyg.2013.00369

Link, T., Huber, S., Nuerk, H.-C., and Moeller, K. (2014). Unbounding the mental number line - new evidence on children's spatial representation of numbers. *Front. Psychol.* 4:1021. doi: 10.3389/fpsyg.2013.01021

Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., and Lindberg, S. (2013). Developmental changes in the association between approximate number representations and addition skills in elementary school children. *Front. Psychol.* 4:783. doi: 10.3389/fpsyg.2013.00783

Lopes-Silva, J. B., Moura, R., Júlio-Costa, A., Haase, V. G., and Wood, G. (2014). Phonemic awareness as a pathway to number transcoding. *Front. Psychol.* 5:13. doi: 10.3389/fpsyg.2014.00013

Mazzocco, M. M. M., Murphy, M. M., Brown, E. C., Rinne, L., and Herold, K. H. (2013). Persistent consequences of atypical early number concepts. *Front. Psychol.* 4:486. doi: 10.3389/fpsyg.2013.00486

Mejias, S., and Schiltz, C. (2013). Estimation abilities of large numerosities in Kindergartners. *Front. Psychol.* 4:518. doi: 10.3389/fpsyg.2013.00518

Mussolin, C., Noël, M.-P., Pesenti, M., Grandin, C., and De Volder, A. G. (2013). Neural correlates of the numerical distance effect in children. *Front. Psychol.* 4:663. doi: 10.3389/fpsyg.2013.00663

Patro, K., Nuerk, H.-C., Cress, U., and Haman, M. (2014). How number-space relationships are assessed before formal schooling: a taxonomy proposal. *Front. Psychol.* 5:419. doi: 10.3389/fpsyg.2014.00419

Van Viersen, S., Slot, E. M., Kroesbergen, E. H., van't Noordende, J. E., and Leseman, P. P. M. (2013). The added value of eye-tracking in diagnosing dyscalculia: a case study. *Front. Psychol.* 4:679. doi: 10.3389/fpsyg.2013.00679

Charting the role of the number line in mathematical development

Jo-Anne LeFevre[1,2]*, *Carolina Jimenez Lira*[1], *Carla Sowinski*[1], *Ozlem Cankaya*[1], *Deepthi Kamawar*[1,2]
and Sheri-Lynn Skwarchuk[3]

[1] Institute of Cognitive Science, Carleton University, Ottawa, ON, Canada
[2] Department of Psychology, Carleton University, Ottawa, ON, Canada
[3] Faculty of Education, University of Winnipeg, Winnipeg, MB, Canada

Individuals who do well in mathematics and science also often have good spatial skills. However, the predictive direction of links between spatial abilities and mathematical learning has not been firmly established, especially for young children. In the present research, we addressed this issue using a sample from a longitudinal data set that spanned 4 years and which includes measures of mathematical performance and various cognitive skills, including spatial ability. Children were tested once in each of 4 years (Time 1, 2, 3, and 4). At Time 3 and 4, 101 children (in Grades 2, 3, or 4 at Time 3) completed mathematical measures including (a) a number line task (0–1000), (b) arithmetic, and (c) number system knowledge. Measures of spatial ability were collected at Time 1, 2, or 3. As expected, spatial ability was correlated with all of the mathematical measures at Time 3 and 4, and predicted growth in number line performance from Time 3 to Time 4. However, spatial ability did not predict growth in either arithmetic or in number system knowledge. Path analyses were used to test whether number line performance at Time 3 was predictive of arithmetic and number system knowledge at Time 4 or whether the reverse patterns were dominant. Contrary to the prediction that the number line is an important causal construct that facilitates learning arithmetic, no evidence was found that number line performance predicted growth in calculation more than calculation predicted number line growth. However, number system knowledge at Time 3 was predictive of number line performance at Time 4, independently of spatial ability. These results provide useful information about which aspects of growth in mathematical performance are (and are not) related to spatial ability and clarify the relations between number line performance and measures of arithmetic and number system knowledge.

Keywords: number line, spatial abilities, number system knowledge, arithmetic, mathematics

INTRODUCTION

Individuals who do well in mathematics and science also usually have good spatial skills (Wai et al., 2009). However, causal links between spatial ability and mathematical learning have not been firmly established, especially for young children (Mix and Cheng, 2012). Thus, despite substantial correlational evidence for links between these two cognitive domains, especially for older children and adults (Mix and Cheng, 2012), a large amount of research remains to be done explaining the nature and direction of the links. In recent research, it has been suggested that a specific skill, that is, estimating the location of numbers on a number line, mediates the link between spatial abilities and growth in conventional mathematical knowledge (Gunderson et al., 2012). The goal of the present research was to use longitudinal data to test the predictive pathways among spatial abilities, number line task performance, and mathematical learning for children in elementary school.

It seems relatively uncontroversial that spatial tasks, especially those tapping visual-spatial working memory or mental rotation, are correlated with mathematical task performance (Mix and Cheng, 2012). One type of mathematical task, the number line task, has received a great deal of attention in this regard. For the number line task, children are shown a line with the left end marked as 0 and the right end marked with a number such as 10, 100, or 1000 (Siegler and Opfer, 2003; Siegler and Booth, 2004; Booth and Siegler, 2006; Laski and Siegler, 2007). In the number-to-position version of the task, children are shown or told a number (e.g., 47) and asked to indicate its location on the number line (Laski and Siegler, 2007). In the position-to-number version, they are asked to estimate the number indicated by a marked position on a line (Siegler and Opfer, 2003; Ashcraft and Moore, 2012). Children above 6 years of age appear to understand the requirements of the task and it is relatively simple and easy to administer. In many studies over the last 10 years, performance on the number line task has been shown to correlate with various standardized mathematical performance measures and with assessments of measurement, numerosity, and computational estimation (Siegler and Booth, 2004; Booth and Siegler, 2006).

The number line task seems ideal for examining the links between spatial abilities and mathematical learning because it

requires knowledge and processes from both domains. Children presumably must be relatively familiar with the number system in the range specified in the particular version of the number line task, and they need to use proportional reasoning skills (or some other strategy) to connect the number to the position on the line (or vice-versa). Ashcraft and Moore (2012), using a position-to-number version of a 0–1000 number line, showed that children from Grades 3 to 5 and adults used an implicit midpoint reference to bisect the line and guide their number choices, and thus, showed very good performance for locations close to 500. The adults and older children also showed good performance (i.e., less variability) for locations close to 250 and 750, suggesting they were using a proportional reasoning strategy of dividing the line into quarters. These data support the view that adults and older children use proportional reasoning strategies to make decisions about locations on the number line (see Barth and Paladino, 2011). For younger children, data collected by Moeller and colleagues (2009) with Grade 1 children and by Bouwmeester and Verkoeijen (2012) with children in kindergarten, grade 1, and grade 2 suggested that many children use a counting-based strategy that results in relatively ordinal and linear performance at the smaller end of the number line, with variable performance for the end, resulting in either logarithmic fits or patterns described by two separate lines. Examples of some patterns of performance are shown in **Figure 1** for children in Grade 2 from the current sample.

A few children do poorly on this task, as shown by the pattern in **Figure 1A**. In contrast, most children at this age show some understanding of the relative position of the numbers on the line. For example, the child whose data are shown in **Figure 1B** produced ordinally-correct positions that are overestimates below 250, and relatively uniform (and thus, non-ordinal) responses for the larger numbers. Depending on the exact shape, this pattern is fit better with logarithmic, quadratic, or exponential functions; however, the linear fit in this case is also statistically significant. Moeller et al. (2009) showed that two different regression lines, one for the numbers in the lower range and one for the numbers in the higher range, also provide a good fit for many children showing this pattern (see Bouwmeester and Verkoeijen, 2012).
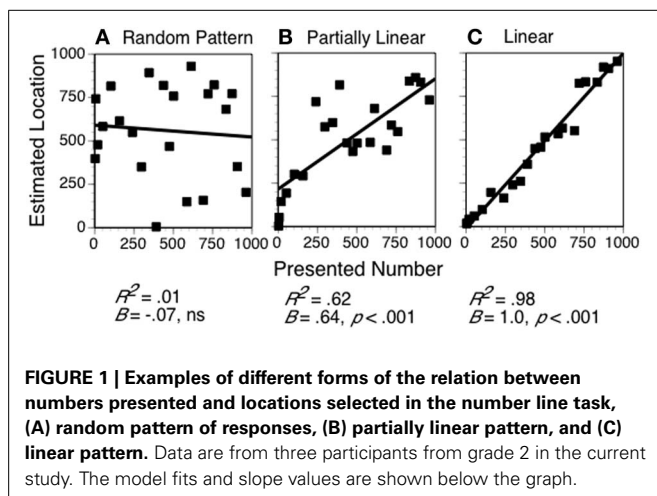
One interpretation of this pattern is that these children have relatively intact number knowledge up to a certain point, but a weak grasp of the larger numbers and their inter-relations. Finally, consider the highly accurate linear pattern shown in **Figure 1C**. Individuals whose performance shows a linear pattern presumably have a strong grasp of the ordinal positions of numbers, including the uniform spacing of the numbers in the indicated range, and sufficient spatial skills to produce very accurate positioning. One goal of the present study was to examine how children's estimates changed over time in relation to their spatial and numerical skills.

Some researchers (Siegler and Opfer, 2003; Siegler and Booth, 2004; Booth and Siegler, 2006; Laski and Siegler, 2007) have interpreted less linear patterns as representative of a non-linear internal mental representation for magnitude rather than a reflection of different strategic processes (cf. Moeller et al., 2009; Bouwmeester and Verkoeijen, 2012). The view of number line performance as an index of children's internal representation of magnitude has not been conclusively proven and is not a necessary assumption: Number line task performance is an interesting and relevant measure even if it does not reflect an internal mental number line. Strong claims about number line performance as a reflection of an internal representation have been based on finding high correlations between number line performance and other mathematical tasks (Booth and Siegler, 2008) or on the increasing linearity of the patterns that children produce as they get older (Booth and Siegler, 2006). More recently, research showing that the strategies children adopt on the task are strongly related to their patterns of performance (Moeller et al., 2009; Bouwmeester and Verkoeijen, 2012) suggest that it is not necessary to postulate direct connections to an internal mental number line to understand performance. Regardless of the interpretation of number line task performance that is assumed, the data show that younger children tend to produce less linear number line estimations than older children. Thus, progress toward linear performance on the number line task can be used as an index of growth in children's understanding of the symbolic number system. Because placement of targets on the number line is a necessary component of the task, we hypothesized that spatial reasoning abilities would be related to growth in number line performance. Importantly, however, we also expected that children's knowledge of the number system would influence their developing number line performance.

We know of only one other study that evaluated changes in number line task performance over time in relation to both spatial skill and mathematically-relevant knowledge. Gunderson et al. (2012; Experiment 1) had children complete a measure of spatial processing (mental rotation) at the beginning of first or second grade ($N = 152$). They also completed a 0–1000 number line task at both time points. At the end of the year, they completed a measure of arithmetic problem solving. Gunderson et al. found that spatial ability predicted growth in number line performance across the year, as did the arithmetic measure. These results were the first to show that spatial ability is linked to improvements in children's number line task performance. In a second study, Gunderson et al. investigated whether number line task performance would predict later mathematical achievement. In

FIGURE 1 | Examples of different forms of the relation between numbers presented and locations selected in the number line task, (A) random pattern of responses, (B) partially linear pattern, and (C) linear pattern. Data are from three participants from grade 2 in the current study. The model fits and slope values are shown below the graph.

this study, 42 children completed spatial measures at age five, a 0–100 number line task at age six, and an approximate symbolic arithmetic task at age eight. Performance on all three measures were correlated, however, the relation between spatial ability and approximate arithmetic was completely mediated by number line task performance. Gunderson et al. suggested that their findings supported causal links between early spatial ability, acquisition of a linear number line, and later number knowledge. Another goal of the present research was to further test this proposed causal chain and to examine whether this finding holds for older children than previously tested.

In the present research, longitudinal data from a large study of children's early mathematics development was used to evaluate three hypotheses. First, we hypothesized that spatial ability would be correlated with number line performance, as well as with traditional measures of mathematical performance (i.e., arithmetic and number system knowledge). Second, we hypothesized that spatial ability would predict growth (change over time) in number line task performance, above and beyond its relation with other mathematical skills. This hypothesis was based both on the findings of Gunderson et al. (2012) and on the assumption that the number line task requires explicit spatial processing in the form of proportional reasoning (Ashcraft and Moore, 2012). Third, we tested the predictive pathways between number line task performance and several other mathematical measures. Using cross-lag panel analysis, we assessed the longitudinal relations between number line task performance and measures of mathematical performance. These analyses provide a stringent test of the hypothesized directional links between number line task performance and (for example) arithmetic, because performance measures were available longitudinally for all measures.

METHODS AND MATERIALS
PARTICIPANTS
The participants were part of a longitudinal study of over 500 children that spanned 4 years (see LeFevre et al., 2010 for analyses of a younger cohort from the same study; also Kamawar et al., 2010; LeFevre et al., 2013). Children were tested once each year (i.e., Time 1, 2, 3, and 4). Data from 101 children (55 girls; 46 boys) who had completed the number line task twice (at Times 3 and 4 of the project) were analyzed in this paper. At Time 3 the participants were in Grade 2, 3, or 4 (ns of 52, 27, and 21). The mean ages were 7:10, 8:10, and 9:10 (in years: months) at Time 3 for children in grades 2, 3, and 4, respectively. Children were recruited from several different schools in two Canadian cities. Eighty-two of the children were monolingual English speakers and the others spoke another language in addition to English. The majority of children came from two parent families and most parents had education beyond high school. Thus, the sample was predominantly middle class.

MEASURES
Children completed a range of cognitive and mathematical measures. More detail about each measure is provided below. In the present analysis, we used measures of (a) spatial ability, (b) number line estimation, (c) number system knowledge, and (d) arithmetic. Control variables included vocabulary, grade, and gender. The mathematical measures were completed by the children twice (Years 3 and 4 of the study). The mathematics measures utilized in this study were chosen to represent typical symbolic number and arithmetic tasks.

Spatial ability
Children were administered the Analogy subtest of the Cognitive Intelligence Test Nonverbal (CIT; Gardner, 2000) either at Time 2 (children from Grade 2 or 4 group) or Time 3 (children from Grade 3 group) of the study. On each trial, children were shown six squares with visual patterns in only five of the squares. They are then asked to pick the "missing pattern" from a selection of patterns at the bottom of the page. This task requires the use of analogical reasoning, mental rotation, and spatial processing for the child to identify the form or design that best completes the pattern. Standard scores on this test have a mean of 100 and a standard deviation of 15. Scaled scores have a mean of 10 and a standard deviation of 3. The test manual reports that the reliability for this subtest was calculated using the Kuder-Richardson formula at each one year age level: for 6, 7, and 8 year-olds reliability was 0.82; for the 9 year-olds the reliability was 0.75. Three children were missing a score so these were replaced with the overall mean for the task.

Children also completed a spatial memory span measure that was administered on a laptop computer. Similar spatial span measures have been used in other studies (e.g., Berch et al., 1998; Rasmussen and Bisanz, 2005). In this task, after the children pressed the "GO" button, a set of nine green circles (lily pads) appeared on the screen and the children watched a cartoon frog as it jumped from one lily pad to another at 1 s intervals. After viewing a sequence of "jumps", the child was given a pointer and asked to reproduce the sequence. As the child pointed to each location on the screen, the experimenter moved the cursor to the corresponding location and clicked on it so the sequence was saved in the computer.

Children completed one practice trial, during which the experimenter watched the frog and then pointed to the two locations in sequence. The test trials consisted of two sequences for each length with the spans increasing in length by one after each pair. The test trials began with two locations and went up to eight. The task ended when the child made errors on two consecutive trials at a specific length. For the analysis, data was used for this task from Time 2; however, Time 1 data were used for 10 children who were missing these data at Time 2. The score was the total number of sequences correct (maximum score of 14). The scores were standardized by creating z-scores by age group. The split-half reliability of this task was 0.78 based on the sum of the first and second trial at each length.

Numeration
Numeration was assessed using the numeration subtest of the Key Math Test-Revised (Connolly, 2000). This Canadian norm-referenced test has two alternate forms; for this study Form B was administered at Time 3 and Form A was administered at Time 4. Children typically attempted between 18 and 30 items on this task.

This task assesses children's knowledge of the number system by having them name numbers and demonstrate understanding of the ordering of symbolic quantities and an understanding of place value for numbers between 100 and 1000. For example, they may be asked to put three numbers in order. Raw scores were used for the analyses; standardized scores on this test have a mean of 10 and a standard deviation of 3. The alternate form reliability coefficient for the numeration subtest within an American sample was 0.75 (Connolly, 2000).

Calculation

Children completed the Calculation subtest of the Woodcock-Johnson Tests of Achievement—Revised (Woodcock and Johnson, 1989). During this paper and pencil test, the children solved mathematical problems such as $3 + 4$ or $15 - 8$ presented in order of increasing difficulty. Testing was stopped after the child incorrectly answered six questions in a row. The score is standardized by grade. Split half reliabilities were reported as 0.93 and 0.89 for children aged 6 and 9, respectively. Children typically attempted between 20 and 32 items on this task.

Number line task

This study implemented a computerized version of the number line task introduced by Siegler and Opfer (2003). The task was described as a game called "Number Line Road." The child was shown a straight line with 0 on its left end and 1000 on its far right end. After the child pressed the "GO" button, a number appeared on the upper right of the screen and the child had to use the mouse to place the cursor—which appeared as a red car with a red straight line beneath it—at the spot where the child estimated the number to be located along the road. When the child clicked on a location the car's last location was retained briefly and then a car horn sounded to indicate a successful placement.

Children were given three practice trials on which they had to place the car on "stop lines" located at 500, 0, and 1000 on the number line road. The test stimuli at Time 3 were those used by Booth and Siegler (2006), and included 22 trials using the following numbers: 3, 7, 19, 52, 103, 158, 240, 297, 346, 391, 438, 475, 502, 586, 613, 690, 721, 760, 835, 874, 907, 962. The order of presentation of the stimuli was randomized separately for each child. The stimuli used at Time 4 were those used by Laski and Siegler (2007). They included 25 trials that were selected to evenly distribute the numbers across the number line. Thus, they included four numbers between 0 and 100, four numbers between 900 and 1000, two numbers from each other decade and distances matched from the endpoints. The numbers used were: 6, 18, 59, 97, 124, 165, 211, 239, 344, 383, 420, 458, 500, 542, 580, 617, 656, 761, 789, 835, 876, 903, 991, 982, 994. A linear regression was run for every child to determine the R^2 and linear slope of the fit between actual and estimated locations. Reliability coefficients for the initial larger sample of children computed based on split-half of odd and even trials at Time 3 and Time 4 were 0.856 and 0.866 (Cronbach's alpha, ns of 203 and 238).

Receptive vocabulary

At Time 1, all children completed the Peabody Picture Vocabulary Test—Third Edition (PPVT III; Dunn and Dunn, 1997) as a measure of their receptive vocabulary. During this test, children were shown a set of four pictures and asked to select the picture that best suits the target word presented by the examiner. This test is norm-referenced and is standardized by age; it has a mean of 100 with a standard deviation of 15. Reliability for this test is reported as 0.94 (Cronbach's alpha).

PROCEDURE

Children were tested during two one-on-one 30-min sessions, or in one 60-min session, that took place within the children's schools. The standardized tests were completed in one session and the computer adapted tests were administered during the other session. Each session was conducted by a different experimenter. The computer adapted tests were administered on a laptop computer. The keyboard was covered except for the spacebar which acted as the "GO" button. In order to have the child's attention on the screen when the stimuli appeared, the child was asked to initiate each trial by pressing the "GO" button.

RESULTS

DESCRIPTIVE STATISTICS

Means, standard deviations, and skew for the measures are shown in **Table 1**. For ease of comparison, the standard scores are shown in the table for the Calculation and Numeration measures; however, raw scores for these variables had a larger range than the standard scores and thus, were used in the correlational and regression analyses.

In many previous studies using the number line task, the performance measure has been the R^2 value for the linear fit between presented numbers and number line locations for each participant. However, as shown in **Table 1**, the linear R^2 value is highly negatively skewed for this age group on a 0–1000 number line. Calculating the arcsine transformation of the R^2 values (Gunderson et al., 2012) reduced the skew somewhat, but it was still substantial. In contrast, although the slope value for the linear fit is still significantly skewed at Time 3, with $z = 2.92$, at Time 4 the skew is reduced and no longer significant. Accordingly, the linear slope value was used in the regression and cross-lag analyses. The slope value approaches 1.0 as linearity increases. Slopes greater than 1 are possible but were infrequent (one at Time 1 of 1.08; eight at Time 2 with the largest 1.09). Thus, slopes were used to index number line performance. They capture both the absolute and relative accuracy of children's number line performance. Linear R^2 could be very high as long as the ordinal positions of the numbers were preserved, but slopes will continue to improve as the locations are placed more accurately (i.e., when children are neither under nor overestimating at the ends of the range). We used slope as the index of number line performance rather than an accuracy measure (e.g., percentage of absolute error) because we wanted the dependent variable to be similar to that used by Gunderson et al. and by Siegler in most of his studies.

For spatial ability, a composite variable was calculated using principal components analysis with Spatial reasoning and Spatial Span. The factor score was used in all subsequent analyses (hereafter referred to as the Spatial Factor). It accounted for 69.8% of the variance and the two measures loaded at 0.83 on the factor.

Correlations among the measures are shown in **Table 2**. Correlations were based on raw scores for the mathematics

Table 1 | Descriptive statistics for predictor and outcome measures.

Measures	Time 1, 2 or 3[a]			Time 4		
	M	*SD*	Skew	*M*	*SD*	Skew
COGNITIVE SKILLS						
Vocabulary	109.9	10.4	−0.223			
Spatial reasoning	104.4	13.8	−0.623**			
Spatial span	6.0	1.9	0.312			
MATHEMATICAL SKILLS						
Numeration scaled	12.8	2.8	−0.240	11.0	3.3	−0.272
Numeration raw	16.5	3.4	−0.091	17.4	3.4	−0.230
Calculation standard	96.6	14.0	−0.078	91.2	14.6	0.307
Calculation raw	14.2	4.0	0.168	16.5	4.7	0.682**
NUMBER LINE						
Linear slope	0.64	0.19	−0.701**	0.74	0.20	−0.275
Linear R^2	0.76	0.21	−1.705***	0.85	0.16	−1.618***
Arcsine linear R^2	2.18	0.53	−1.551***	2.44	0.43	−1.129***

Vocabulary, spatial reasoning, and calculation are standard scores with means of 100 and standard deviations of 15. Numeration scores are standardized with a mean of 10 and a standard deviation of 3.

**p < 0.01; ***p < 0.001.

[a]Cognitive skills were assessed either at Time 1 (Vocabulary), Time 2 (Spatial span for all children, Spatial reasoning for youngest and oldest groups) or Time 3 (Spatial reasoning for middle group) whereas all reported mathematical skills were assessed at Time 3 and Time 4.

Table 2 | Correlations among predictors and outcomes; simple correlations below the diagonal; partial correlations above the diagonal (controlling for sex, grade, and vocabulary).

	Control variables			Spatial factor	Number line		Numeration		Calculation	
	Grade	Sex	Voc		T3	T4	T3	T4	T3	T4
Sex	−0.08									
Vocabulary	0.07	0.25*								
Spatial factor	0.12	0.01	0.24*		0.27**	0.43**	0.39**	0.32**	0.38**	0.24*
Number line T3	0.44**	0.29**	0.27**	0.31**		0.56**	0.25*	0.17	0.31**	0.28**
Number line T4	0.43**	0.36**	0.27**	0.43**	0.71**		0.48**	0.32**	0.41**	0.35**
Numeration T3	0.64**	0.15	0.37**	0.41**	0.53**	0.66**		0.52**	0.42**	0.36**
Numeration T4	0.49**	0.14	0.24*	0.36**	0.42**	0.52**	0.70**		0.38**	0.47**
Calculation T3	0.70**	0.00	0.13	0.36**	0.51**	0.57**	0.68**	0.59**		0.52**
Calculation T4	0.72**	0.13	0.14	0.26*	0.54**	0.58**	0.68**	0.66**	0.76**	

*p < 0.05; **p < 0.01.

measures because the range and variability was greater for raw than standard scores. Thus, grade was correlated with all of the unstandardized measures, as expected. For all measures (number line, Numeration, and Calculation), older children had better performance than younger children. Boys had higher scores on the number line measures than girls. Vocabulary was correlated with all of the other measures except for calculation. Therefore, grade, gender, and vocabulary were controlled in the analyses of number line task performance (partial correlations are shown above the diagonal). Performance was correlated at 0.70 or higher from 1 year to the next for the number line, Numeration, and Calculation measures, suggesting considerable stability of these measures. The control variables generally moderated the correlations. Because there was not enough data to evaluate patterns of performance for each grade separately, analyses were done across grades and grade was included as a control variable.

PATH ANALYSES

The longitudinal relations among spatial ability, number line performance, and calculation were evaluated using simultaneous path analysis in Mplus (Version 6; Muthén and and Muthén, 1998/2011). The β values fit by the model, including all significant direct effects, are shown in **Figure 2**. The significant indirect effects are listed in **Table 3**. Significance of the indirect effects was tested using 95% confidence intervals calculated using bias-corrected bootstrap sampling (Geiser, 2013). Although it is not shown in the figure, the regressions controlled for grade and sex for number line and for grade for Calculation.

As shown in **Figure 2**, our data replicated the concurrent and longitudinal relations among spatial ability and number line performance reported by Gunderson et al. (2012), specifically, that growth in number line performance from Time 3 to Time 4 was predicted by spatial ability (i.e., the direct effect from spatial

FIGURE 2 | Path analysis showing longitudinal relations among spatial ability, number line performance, and Calculation ($R^2 = 0.356$, and 0.584 for number line at T3 and T4, respectively; $R^2 = 0.563$ and 0.611 for Calculation at T3 and T4, respectively). $**p < 0.01$.

FIGURE 3 | Path analysis showing longitudinal relations among spatial ability, number line performance, and Numeration ($R^2 = 0.351$, and 0.617 for number line at T3 and T4, respectively; $R^2 = 0.578$ and 0.482 for Numeration at T3 and T4, respectively). The residual variances between number line and Numeration were not significant at either Time 3 or Time 4. $*p < 0.05$; $**p < 0.01$.

Table 3 | Significant effects (standardized) of spatial ability on number line and calculation at Time 4.

	β	Confidence intervals[a]	
		Lower 2.5%	Upper 2.5%
SPATIAL TO NUMBER LINE TIME 4			
(a) Total effect	0.389	0.265	0.514
(b) Direct effect	0.188	0.058	0.318
(c) Indirect through number line at T3	0.138	0.047	0.230
(d) Indirect through calculation at T3	0.063	0.011	0.115
SPATIAL TO CALCULATION TIME 4			
(a) Total effect	0.240	0.147	0.332
(b) Indirect through number line T3	0.052	0.006	0.097
(c) Indirect through calculation T3	0.188	0.106	0.269

There was no direct effect from spatial ability to calculation at Time 4.
[a]Confidence intervals were calculated with bias-corrected bootstrapping in Mplus (10,000 samples).

ability to number line at Time 4). There were also significant indirect effects of spatial ability on number line and Calculation performance at Time 4, mediated through Calculation and number line at Time 3 (see **Table 3**).

These findings extend the results reported by Gunderson et al. (2012) to a slightly older group of children and provide a more complete picture of the longitudinal relations between number line and Calculation because they model the autoregressive effects for both of these variables. Although these results are consistent with Gunderson et al.'s reported pattern of results, the more complete picture shown in the current analysis does not support their larger conjecture that number line knowledge is the critical causal variable that is most relevant for understanding how spatial ability is linked to the development of mathematical skills. The indirect effect of spatial ability on Calculation at Time 4 that is mediated through Calculation at Time 3 (0.188) is significantly larger than the indirect effect mediated through number line performance at Time 3 (0.052, confidence intervals do not overlap). This pattern suggests that it is important to consider a broader model of how spatial abilities may influence the development of mathematical skills.
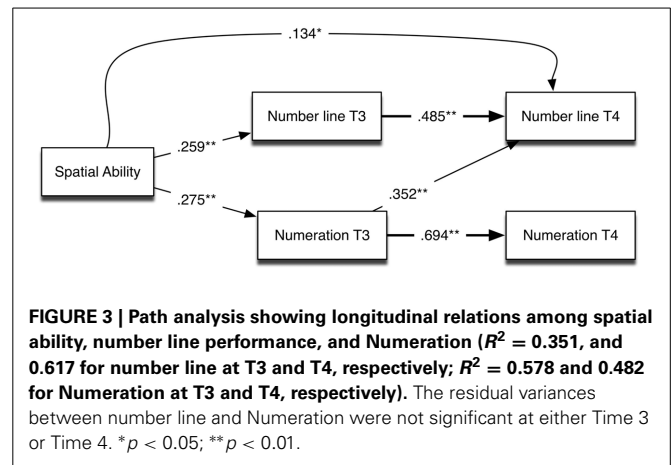
We further tested the possibility that the influence of spatial abilities on later mathematical skills is mediated through skills other than number line performance by evaluating and testing a model that included number line and Numeration knowledge (see **Figure 3**). In this model, control variables included sex and grade for number line at Time 3 and vocabulary and grade for Numeration at Time 3. Note that the Numeration measure, in comparison to the number line task, is a broader assessment of children's understanding of the symbolic number system in the thousands and beyond. As shown in **Figure 3**, spatial ability was a significant predictor of both number line and Numeration performance at Time 3, and showed a direct link to number line performance at Time 4. In this model, however, it is also clear that the indirect influence of spatial ability on Numeration at Time 4 was mediated only through Numeration at Time 3, because the cross-lagged path through number line at Time 3 was not significant. As shown in **Table 4**, the indirect effect of spatial ability on growth in Numeration at Time 4 (i.e., mediated by Numeration at Time 3) was significant. Finally, the indirect (mediated) path from spatial ability through Numeration at Time 3 to the number line at Time 4 was significant (see **Table 4** for indirect effects), supporting the conclusion that spatial ability is related to the development of mathematical skills via multiple pathways. This model also shows that Numeration, as a measure of number system knowledge, predicts growth in children's number line performance.

DISCUSSION

How are spatial abilities related to children's mathematical learning? In support of our first hypothesis that spatial abilities are related to children's performance on various mathematical tasks, we found significant correlations between spatial ability and number line task performance, arithmetic, and number system knowledge (Booth and Siegler, 2006; Lachance and Mazzocco, 2006; Gunderson et al., 2012). We also found support for our second hypothesis, that spatial ability predicts growth in number line knowledge. These results were consistent with the results of Gunderson et al. (2012) and extend the link between spatial ability and mathematics to somewhat older children. Performance on the number line presumably requires spatial abilities because, even when children understand the number system range, they

Table 4 | Significant effects (standardized) of spatial ability on number line and numeration at Time 4.

	β	Confidence intervals[a]	
		Lower 2.5%	Upper 2.5%
SPATIAL TO NUMBER LINE TIME 4			
(a) Total effect	0.356	0.233	0.479
(b) Direct effect	0.134	0.009	0.258
(c) Indirect through number line T3	0.126	0.041	0.210
(d) Indirect through numeration T3	0.097	0.030	0.164
SPATIAL TO NUMERATION TIME 4			
(a) Indirect through numeration T3	0.191	0.088	0.293

There was no direct effect from spatial ability to numeration at Time 4 and no indirect effect through number line at Time 3 and thus, the total effect is indirect through numeration.

[a]*Confidence intervals were calculated with bias-corrected bootstrapping in Mplus (10,000 samples).*

still need to determine the approximate location of a number along a continuum and indicate it as an explicit spatial location.

We did not find support for our third hypothesis, that the number line task is specifically predictive of growth in arithmetic knowledge. Gunderson et al. (2012) found that the link between spatial ability (age 5) and symbolic approximate arithmetic (age 8) was mediated by number line task performance. Although the pattern observed by Gunderson et al. was also present in our data, such that the influence of spatial ability on Calculation is mediated through number line performance (i.e., an indirect effect as shown in **Table 3**), there was also an indirect effect of spatial ability on Calculation at Time 4 that was mediated through Calculation at Time 3 and the latter effect was larger than the former. In other words, number line task performance does not have a privileged role in linking spatial ability to mathematical learning. Instead, we see in **Figure 2** that the cross lagged relations between number line and Calculation performance are significant in both directions and of a similar size. These results have multiple interpretations, including (a) the two tasks required overlapping knowledge or skills (e.g., spatial abilities), and/or (b) both are related to performance on some other unmeasured variable or variables. Thus, the present data do not support the strong claim for a central role for number line knowledge in the development of other mathematical skills (see also Sasanguie et al., 2013; cf. Booth and Siegler, 2008; Gunderson et al., 2012).

More generally, in combination with the results of the longitudinal analysis of number line and Numeration knowledge, our results support the view that the number line task is a complex measure that improves with the development of a variety of relevant mathematical and spatial skills. In the cross-lag analysis, number system knowledge was directionally linked to growth in number line task performance. This pattern suggests that understanding of the number system in a range specified in the tested version of the number line task drives improvements in number line task performance (Sasanguie et al., 2013). Ebersbach et al. (2008) also reported that children (5–9 years old) perform linearly on the number line task when the target numbers are those within their counting range. The Numeration measure used in the present research indexes children's grasp of place value structure between 100 and 1000, and thus, reflects the requisite number system knowledge required to perform well on the number line to 1000. The lack of indirect effects from number line performance at Time 3 to Numeration at Time 4 is consistent with the assumption that the Numeration measure indexes number system knowledge that is necessary to perform the number line task, rather than the reverse.

The present findings emphasize that the number line task requires both spatial abilities and number system knowledge (in the range specified by the endpoints of the number line). Children in grades 2 through 5 are still developing their number knowledge in the range to 1000. Many of the children showed a pattern of performance (see **Figure 1B**), which suggests that their understanding of numbers did not extend past the low hundreds. This pattern is similar to the finding with younger children that their understanding of cardinality does not generalize beyond the number of objects they can count until they gain an understanding of the way in which counting can be used to determine the size of a set (Wynn, 1992; Sarnecka and Carey, 2008). Similarly, children's number naming performance grows gradually as they work with numbers within a certain range (Skwarchuk and Anglin, 2002). In summary, the present data emphasize the important link between number system understanding and the linearity of number line performance.

Given the causal model proposed for the relation between number line performance and Calculation, our finding of joint rather than directional links suggests caution in drawing conclusions about number line task performance as a reflection of children's underlying numerical representations. The overlapping relations between Calculation and number line performance could also reflect the mutual influence that children's conceptual knowledge (i.e., understanding how the number system works in this case indexed by performance on the number line) and specific procedural skills (i.e., the steps that should be taken to solve a mathematical problem) have on each other over the course of the development of their ability to understand and solve arithmetical tasks (Rittle-Johnson et al., 2001; Gilmore, 2006). Booth and Siegler (2008) argued that children's number line task performance reflects their representation of quantity and thus, should be influential in the development of arithmetic skills. They found support for this view when the number line was used in an intervention to represent quantities and to model addition. Kucian et al. (2011) also found transfer from a number line training condition to arithmetic, however, as in the Booth and Siegler intervention, the training condition included both number line and arithmetic practice. Without the explicit training on using the number line as an arithmetic tool, children may not connect number line knowledge to calculation and thus, there may not be a causal link between the two aspects of mathematical knowledge. Instead, the current research suggests that the number line task indexes children's understanding of the number system, and in particular, the ordinal relations among symbolic representations. Other research indicates that number line task performance also reflects children's ability to use proportional reasoning strategies to map these symbolic representations accurately to a specific physical extent, perhaps part of the link with spatial reasoning.

Calculation presumably also requires some or all of the same skills, and thus, the two tasks show improvements that are related but are not explicitly directional. Some caution is recommended, therefore, in training number line task performance with the expectation that it will transfer to improved calculation skills. The present results suggest a need for a better understanding of how number line training might provide benefits that are independent of training in specific fundamental skills.

How do the present data advance our understanding of the relations between spatial and mathematical abilities? First, the finding that spatial abilities predict growth in number line task performance replicates and extends the findings of Gunderson et al. (2012) to older children. These results are consistent with the view that spatial abilities are one of the precursor cognitive skills that support children's learning of related mathematical constructs (LeFevre et al., 2010; Mix and Cheng, 2012). The number line task has obvious spatial processing requirements in that the child has to align the numbers according to their place value within two predetermined endpoints of a continuum. Second, our finding that number system knowledge is also a predictor of growth in number line task performance supports a view of the number line task as an index of children's understanding of the number system in a specified range. As shown by Thompson and Opfer (2010), children can show very strong linearity for a number line task in a familiar range, and yet show relatively poor performance on a number line task in an unfamiliar range. To the extent that they use strategies that involve creating an implicit midpoint reference (e.g., at 500 for a 0–1000 number line; Ashcraft and Moore, 2012) and/or apply a proportional reasoning strategy (e.g., 114 is about 10% of 1000, so it is about 1/10th of the distance from the left; Barth and Paladino, 2011), both number system understanding and spatial reasoning are required when children develop strategies and implement resulting procedures to perform the number line task. In summary, it is important to stress that these findings indicate that spatial knowledge is necessary but not sufficient; growth in number line performance is also driven by earlier knowledge of the number system.

To what extent do these data address the issue of whether the number line task reflects children's use of an internal mental representation of number? Our results are neutral on this point as it is not necessary to postulate a specific internal mental representation to understand children's performance on the number line task. It is more parsimonious to assume that the child's task strategy is reflected in the pattern produced (Bouwmeester and Verkoeijen, 2012). Adults show logarithmic patterns in estimation tasks that include large non-symbolic quantities and linear patterns in estimation tasks with symbolic numbers or small non-symbolic quantities (Dehaene et al., 2008). Presumably they are using their conceptual understanding of how the number system is constructed (i.e., understanding of the base-10 structure, that numbers of equal distance are equally spaced, etc.), in combination with proportional reasoning (Barth and Paladino, 2011; Ashcraft and Moore, 2012) to construct a strategy that is suitable for the particular number line with which they are confronted. Children whose initial knowledge about the number system is limited will show increasingly linear patterns of responding as they gain understanding of the number system in the specified

range, and as they more skillfully apply their spatial knowledge to construct an appropriate strategy.

Although the cross-lagged correlational analysis has some limitations, it is nevertheless, more stringent that using regression to test for possible directional links over time. One limitation of this method is that performance on both tasks may be related to growth in other skills that were not measured in the current research (i.e., causality may be linked to other variables). Thus, we cannot reject the possibility that future research may show stronger pathways. Nevertheless, the pattern of correlations that was observed between calculation and the number line task did not support the strong hypothesis that number line performance is causally linked to calculation (cf. Gunderson et al., 2012). The view that the number line task is an outcome of children's growing number knowledge, rather than a predictor of it, needs further consideration.

An important methodological issue for future research is the assessment of spatial abilities. The measure of spatial ability in the present study was based on two different tasks (spatial reasoning and spatial memory span) and thus, is presumably better than using a single predictor. However, as noted by Mix and Cheng (2012), further theoretical and empirical work on the construct of "spatial ability" will be necessary to adequately test the various possible links between those abilities and mathematical learning and development. Recent research has identified at least three aspects of spatial abilities that may be important including mental rotation, spatial visualization, and disembedding (i.e., the ability to identify target figures in a distracting background), each of which have found to be correlated with different aspects of mathematical development (Mix and Cheng, 2012). The measures used in the present research reflected these abilities in various degrees (e.g., the spatial reasoning task required some rotation and disembedding and the spatial memory task also required visualization). Other spatial abilities may also be involved in these tasks, and may also be involved in various aspects of mathematics. Future research should address the unique relations of various spatial measures to the development of mathematical skills in young children.

In summary, these findings suggest that growth in performance on the number line task reflects children's knowledge of the number system in the specified range in combination with their ability to apply their spatial abilities to create a successful strategy to solve the task. In contrast to the claims of several other researchers, improvements in number line performance did not appear to be causally linked to improvements in other mathematical skills (cf. Booth and Siegler, 2008; Kucian et al., 2011; Gunderson et al., 2012). Although the present research was not designed to directly investigate the internal representations that might be activated when children perform the number line task, other studies suggest that it is not necessary to assume anything about an internal representation to understand the development of number line task performance (Ebersbach et al., 2008; Barth and Paladino, 2011; Ashcraft and Moore, 2012; Bouwmeester and Verkoeijen, 2012). Thus, it may not be necessary to view performance on the number line task as the reflection of an organized internal knowledge structure that is causally linked to further learning or to categorize number line performance as an index

of a fundamental "number sense" (i.e., an internal logarithmic number line). Instead, it may be more useful to view the number line task as a measure of children's ability to skillfully assemble an array of relevant knowledge to perform a complex and (often) novel numerically-relevant task.

REFERENCES

Ashcraft, M. H., and Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *J. Exp. Child Psychol.* 111, 246–267. doi: 10.1016/j.jecp.2011.08.005

Barth, H. C., and Paladino, A. M. (2011). The development of numerical estimation: evidence against a representational shift. *Dev. Sci.* 14, 125–135. doi: 10.1111/j.1467-7687.2010.00962.x

Berch, D. B., Krikorian, R., and Huha, E. M. (1998). The corsi block-tapping task: methodological and theoretical considerations. *Brain Cogn.* 38, 317–338. doi: 10.1006/brcg.1998.1039

Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 41, 189–201. doi: 10.1037/0012-1649.41.6.189

Booth, J. L., and Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Dev.* 79, 1016–1031. doi: 10.1111/j.1467-8624.2008.01173.x

Bouwmeester, S., and Verkoeijen, P. P. J. L. (2012). Multiple representations in number line estimation: a developmental shift or classes of representations? *Cogn. Instr.* 30, 246–260. doi: 10.1080/07370008.2012.689384

Connolly, A. J. (2000). *KeyMath Revised/Updated Canadian Norms.* Richmond Hill, ON: PsyCan.

Dehaene, S., Véronique, I., Elizabeth, S., and Pierre, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science* 320, 1217–1220. doi: 10.1126/science.1156540

Dunn, L. M., and Dunn, L. M. (1997). *Peabody Picture Vocabulary Test-III.* Circle Pines, MN: American Guidance Service.

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., and Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year old children: evidence for a segmented linear model. *J. Exp. Child Psychol.* 99, 1–17. doi: 10.1016/j.jecp.2007.08.006

Gardner, M. F. (2000). *Cognitive (Intelligence) Test: Nonverbal.* Hydesville, CA: Psychological and Educational Publications, Inc.

Geiser, C. (2013). *Data Analysis with Mplus.* New York, NY: Guilford Press.

Gilmore, C. (2006). Profiles of understanding and profiles of development in early arithmetic. *Proc. Br. Soc. Res. Learn. Math.* 26, 7–12.

Gunderson, E. A., Ramirez, G., Beilock, S. L., and Levine, S. C. (2012). The relation between spatial skill and early number knowledge: the role of the linear number line. *Dev. Psychol.* 48, 1229–1241. doi: 10.1037/a0027433

Kamawar, D., LeFevre, J. A., Bisanz, J., Fast, L., Skwarchuk, S. L., Smith-Chant, B., et al. (2010). Knowledge of counting principles: how relevant is order irrelevance? *J. Exp. Child Psychol.* 105, 138–145. doi: 10.1016/j.jecp.2009.08.004

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., et al. (2011). Mental number line training in children with developmental calculia. *Neuroimage* 57, 782–795. doi: 10.1016/j.neuroimage.2011.01.070

Lachance, J. A., and Mazzocco, M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learn. Individ. Differ.* 16, 195–216. doi: 10.1016/j.lindif.2005.12.001

Laski, E. V., and Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Dev.* 78, 1723–1743. doi: 10.1111/j.1467-8624.2007.01087.x

LeFevre, J. A., Berrigan, L., Vendetti, C., Kamawar, D., Bisanz, J., Skwarchuk, S. L., et al. (2013). The role of executive attention in the acquisition of mathematical skills for children in grades 2 through 4. *J. Exp. Child Psychol.* 114, 243–261. doi: 10.1016/j.jecp.2012.10.005

LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: longitudinal predictors of performance. *Child Dev.* 81, 1753–1767. doi: 10.1111/j.1467-8624.2010.01508.x

Mix, K. S., and Cheng, Y. L. (2012). "The relation between space and math: developmental and educational implications," in *Advances in Child Development and Behavior,* Vol. 42, ed J. B. Benson (Burlington,VT: Academic Press), 197–243.

Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H. C. (2009). Children's early mental number line: logarithmic or decomposed linear? *J. Exp. Child Psychol.* 103, 503–515. doi: 10.1016/j.jecp.2009.02.006

Muthén, L. K., and Muthén, B. O. (1998/2011). *MPlus User's Guide, 6th Edn.* Los Angeles, CA: Muthén and Muthén.

Rasmussen, C., and Bisanz, J. (2005). Representation and working memory in early arithmetic. *J. Exp. Child Psychol.* 91, 137–157. doi: 10.1016/j.jecp.2005.01.004

Rittle-Johnson, B., Siegler, R., and Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: an iterative process. *J. Educ. Psychol.* 93, 346–362. doi: 10.1037/0022-0663.93.2.346

Sarnecka, B. W., and Carey, S. (2008). How counting represents number: what children must learn when they learn it. *Cognition* 108, 662–674. doi: 10.1016/j.cognition.2008.05.007

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number—space mappings: what underlies mathematics achievement? *J. Exp. Child Psychol.* 114, 418–31. doi: 10.1016/j.jecp.2012.10.012

Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi: 10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–243. doi: 10.1111/1467-9280.02438

Skwarchuk, S. L., and Anglin, J. M. (2002). Children's acquisition of the english cardinal number words: a special case of vocabulary development. *J. Educ. Psychol.* 94, 107–125. doi: 10.1037/0022-0663.94.1.107

Thompson, C. A., and Opfer, J. E. (2010). How 15 hundred is like 15 cherries: effect of progressive alignment on representational changes in numerical cognition. *Child Dev.* 81, 1768–1786. doi: 10.1111/j.1467-8624.2010.01509.x

Wai, J., Lubinski, D., and Benbow, C. P. (2009). Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J. Educ. Psychol.* 101, 817–835. doi: 10.1037/a0016127

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

Woodcock, R. W., and Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery-Revised.* Allen, TX: DLM Teaching Resources.

Severity of specific language impairment predicts delayed development in number skills

Kevin Durkin[1]*, Pearl L. H. Mok[2] and Gina Conti-Ramsden[2]

[1] School of Psychological Sciences and Health, University of Strathclyde, Glasgow, UK
[2] School of Psychological Sciences, Communication and Deafness, The University of Manchester, Manchester, UK

The extent to which mathematical development is dependent upon language is controversial. This longitudinal study investigates the role of language ability in children's development of number skills. Participants were 229 children with specific language impairment (SLI) who were assessed initially at age 7 and again 1 year later. All participants completed measures of psycholinguistic development (expressive and receptive), performance IQ, and the Basic Number Skills subtest of the British Ability Scales. Number skills data for this sample were compared with normative population data. Consistent with predictions that language impairment would impact on numerical development, average standard scores were more than 1 SD below the population mean at both ages. Although the children showed improvements in raw scores at the second wave of the study, the discrepancy between their scores and the population data nonetheless increased over time. Regression analyses showed that, after controlling for the effect of PIQ, language skills explained an additional 19 and 17% of the variance in number skills for ages 7 and 8, respectively. Furthermore, logistic regression analyses revealed that less improvement in the child's language ability over the course of the year was associated with a greater odds of a drop in performance in basic number skills from 7 to 8 years. The results are discussed in relation to the interaction of linguistic and cognitive factors in numerical development and the implications for mathematical education.

Keywords: number skills, number development, specific language impairment (SLI), cognitive factors, linguistic abilities

INTRODUCTION

Developing competence with number is fundamental to everyday life and to education. One of the best predictors of subsequent academic attainment is children's level of rudimentary number skills at school entry (Duncan et al., 2007). Understanding how children deal with this domain and what factors are associated with their learning are topics of theoretical and practical significance. Although most scientists in this area agree that developing number skills depends on complex neural, cognitive, linguistic and interpersonal abilities, the relative contributions of different capacities, and their interrelationships, are controversial and difficult to disentangle. Many researchers have argued that linguistic or linguistically-supported processes, such as labeling, articulating and scaffolding, are critical to mastering basic number, and that conceptual knowledge of the domain is organized, memorized, shared and built upon via language (Durkin, 1993; Wiese, 2003; Carey, 2004; Musolino, 2004; Ginsburg, 2009; Mix, 2009; Negen and Sarnecka, 2012). In contrast, others argue that numerical cognition has phylogenetic and ontogenetic origins independent of the language capacity (Gelman and Butterworth, 2005; Nieder and Dehaene, 2009).

Children with specific language impairment (SLI) are of particular interest in this context (Fazio, 1996, 1999; Arvedson, 2002; Cowan et al., 2005; Donlan et al., 2007; Kleemans et al., 2011, 2012; Nys et al., 2013). Individuals with SLI fall within the normal range of cognitive abilities, show no evidence of neurological damage or hearing impairment yet, relative to peer norms, have deficits in either or both of expression and comprehension of language (Bishop, 1997; Conti-Ramsden and Durkin, 2011). SLI is a common condition affecting some 7% of 5-year-old children (Tomblin et al., 1997). Children with this condition provide a naturally-occurring test of to what extent number skills development is possible when language abilities are compromised but cognitive abilities are within the typical range. In this report, we examine the longitudinal relationship between language ability and number skills development in children with SLI. To explain the background to the study, first we summarize evidence on the relationship between language and number development in the preschool and early school years, and then we describe previous research on number development in children with SLI.

Some important number-related skills emerge before language. Preverbal human infants can discriminate among numbers (Strauss and Curtis, 1981; Wynn, 1992; Cordes and Brannon, 2009), though the basis on which they do so is controversial (Ansari and Karmiloff-Smith, 2002; Sophian, 2007; Izard et al., 2009; Núñez, 2011). While early perceptual abilities may provide a foundation on which to map some number terms, subsequent developments are—for typically developing (TD) children— inevitably interwoven with linguistic experiences and language

development (Spelke and Tsivkin, 2001). Number words are common in everyday parental input to infants and toddlers (Durkin et al., 1986; Fuson, 1988; Bloom and Wynn, 1997; Tare et al., 2008). Furthermore, variation in the extent and quality of number talk in parental speech has been found to predict developments among pre-schoolers in conceptual number knowledge, such as cardinal meanings (Linnell and Fluck, 2001; Levine et al., 2010; Mix et al., 2012).

Developments in language skills are clearly associated with developments in numerical and mathematical abilities. Negen and Sarnecka (2012) showed that the size of a child's general nominal vocabulary (both expressive and receptive) is positively associated with number-concept acquisition, suggesting that noun learning could assist children in discovering the meanings of number words. For example, to understand the number word in a caregiver's utterance such as "Look! There's a mommy duck with three baby ducks!" (Negen and Sarnecka, 2012, p. 2020), it helps if the child already knows what the noun *duck* means and how one creates plural forms. O'Neill et al. (2004) found that measures of narrative competence at 3- to 4- years were predictive of mathematics ability assessed 2 years later.

As children begin to use number words with increasing accuracy in everyday settings, advances follow in their abilities to apply number concepts in contexts where they are trained or tested on number understanding (Mix et al., 2005; Palmer and Baroody, 2011; Reikerås et al., 2012). When children participate in educational contexts, number and early mathematical concepts receive more, and increasingly deliberate, attention (Ginsburg, 2009). Much of this is mediated through language: the verbal instructions of the teacher, the vocabulary and syntax of texts and other school materials (Durkin and Shire, 1987, 1991; Ellerton and Clarkson, 1996; Ginsburg, 2009). Cross-linguistic investigations of number skills development show that children's performances can be helped or hindered by particular features of the way a given language encodes number (Seron and Fayol, 1994; Roberts and Gathercole, 2006; Salehuddin and Winskel, 2009; Zuber et al., 2009; Helmreich et al., 2011; Pixner et al., 2011).

With development, it is increasingly difficult to determine which variables affect progress in number and mathematical abilities, and in practice most children using their number skills will require both computational and linguistic application. How, then, do children whose linguistic abilities are impaired handle this domain?

Children who have difficulties in using language may face particular challenges in advancing their numerical understanding. Indeed, children with SLI face challenges in many areas of development and education (Bishop and Adams, 1990; Conti-Ramsden et al., 2009; Durkin et al., 2009; St. Clair et al., 2010). Difficulties in decoding others' language, in formulating and producing utterances, and in processing textual materials pose considerable burdens when dealing with novel concepts and problem-solving procedures. As Gelman and Butterworth (2005) comment, it would be surprising if there were no effects of language on numerical cognition, and even proponents of independent origins of numerical ability acknowledge that language facilitates the use of numerical concepts.

Previous research with children with SLI confirms that they do lag behind typical peers in progress in number and mathematical abilities, though this depends to some extent on which abilities are tested. Fazio (1994, 1996, 1999) followed a small group of children with SLI from preschool to age 9 years. She found that the children were late in acquiring counting skills and, although they did gradually develop these, they continued to experience difficulties with related rote memory tasks (such as multiplication tables), which she interpreted as indicating that storage or retrieval problems underlie the children's difficulties in this domain. Nelson et al. (2011) found that 4-year-olds with severe language delay scored 1 *SD* below norms on a maths test measuring counting skills and simple addition and subtraction. Kleemans et al. (2011, 2012) showed that naming speed (the ability to retrieve linguistic information from long-term memory) was associated with early verbal numeracy skills in children with SLI (but not in TD children). On the other hand, naming speed did not predict performance on numerical estimation tasks (identifying the location of a number on a number line), which are less verbal. Arvedson (2002) found that on some number tasks designed to minimize verbal processing, 3- to 5-year-old children with SLI performed as well as age matched TD peers, and better than language-matched younger children. However, on tasks that required more verbal processing, and particularly when children were encouraged to count, the participants with SLI performed less well than their TD peers. Donlan et al. (2007), working with 8-year-olds with SLI, reported further evidence of severe deficits in both counting tasks and calculation tasks, compared to TD children. Participants with SLI did not differ from age-matched TD children, though, on a task of arithmetical principles in which judgments were required of the correctness or otherwise of abstract symbolic expressions, the most complex but arguably least linguistic of the tasks that they administered.

In one of the largest studies of number skills in primary school aged children with SLI (aged 8 years), Cowan et al. (2005) found that these participants performed worse than age-matched TD comparison children on a range of tasks, including counting, knowledge of addition combinations, basic calculation, story problems, transcoding (reading and writing multi-digit numbers), and relative magnitude judgments. That is, the disadvantage appeared to be pervasive across number tasks. Zero order correlations revealed consistently stronger associations between number task performances and language comprehension than those found for number task performance and either of working memory and non-verbal reasoning. In standard multiple regressions, language comprehension was the best predictor of variation on most of the number tasks. However, Cowan et al. included all participants, including those with typical development, in their regression analyses. While this provides a good test of the importance of language ability across a range of children, it does not examine directly the impact of severity of impairment among those diagnosed as having SLI.

Taken together, previous research demonstrates that children with SLI are at a disadvantage when it comes to working with number. This is evident from their early difficulties with counting, through to a range of other number-related tasks in childhood. The findings are consistent with the assumption that

language ability bears in some way on number development in the preschool to early school years. However, most of the available evidence is based on comparisons between children with SLI and TD peers at a particular age point. We lack information on the extent to which language impairment predicts progress in number tasks over time. Kleemans et al. (2012) provide valuable longitudinal data to show that grammatical ability at age 6 does contribute toward the prediction of number skills at age 7, but their study did not measure number skills at both age points, leaving the question of what underpins progress in need of further attention.

In this study, we examined a large sample of children with SLI, collecting their raw and standard scores on a standardized number skills test at two age points (7 and 8 years). We measured their psycholinguistic profiles and non-verbal performance IQ (PIQ) at the outset. We investigated changes in number skills over the period tested. Children receive instruction in number at school, and various outcomes are possible. One, which we will describe as the "separate origins" hypothesis, is that, as operating on numbers involves cognitive processes that are independent of language, continuing progress in number performance should be independent of the severity of impairment; this is not to say that children with language impairments will necessarily display exceptionally strong number skills if they have not done so before, but some improvements in raw scores, and some variations in standard scores, should follow from maturational and educational effects and these should not be predictable on the basis of language abilities. An alternative possibility, which we will call the "language contingent" hypothesis, is that the development of number skills in childhood is influenced by the child's linguistic abilities. Thus, on this account, children with SLI should remain at a disadvantage, reflecting a general lagging behind associated with language delay (cf. Conti-Ramsden et al., 2012). Modest improvements in raw scores could be expected, as the children do improve their language skills albeit at a slow pace, and they are exposed to number education; but, as they find it increasingly difficult to keep up with the language demands associated with number work, standard scores should fall relative to peer norms. The present study was designed to test these competing predictions.

MATERIALS AND METHODS
PARTICIPANTS
The participants in this study were originally part of a wider study: the Manchester Language Study (Conti-Ramsden et al., 1997; Conti-Ramsden and Botting, 1999). The original cohort of 242 children, which consisted of 186 boys (76.9%) and 56 girls (23.1%), was recruited from 118 language units in England in 1995, and represented a random sample of 50% of all Year 2 children (approximately 7 years of age) attending language units for at least half of the school week. Language units are specialized classes for children who have been identified with primary speech and language difficulties; the units are usually attached to mainstream schools. Children were excluded from the study if they were reported by their teachers as having frank neurological difficulties, a diagnosis of autism, hearing impairment or a general learning disability. No additional criteria of SLI were used in the selection. Assessment of the children at age 7 after recruitment

into the study indicated that the majority met the traditional criteria of SLI, i.e., the standard scores for at least one test of language was less than 1 SD below the mean whilst PIQ was within 1 SD of the mean, or there was a discrepancy of 40 percentiles or more between the language test scores and PIQ (Conti-Ramsden and Botting, 1999).

A total of 232 of these children were followed up a year later at age 8. For the purpose of this study, only those for whom there were data on their basic number skills measured at both ages have been included in the analyses. This resulted in a total of 229 children, 176 (76.9%) of whom were boys. The mean age of the participants was 7:1 years (6:6–7:9) at wave 1 and 8:1 years (7:5–8:9) at wave 2. The psycholinguistic profiles of the participants at ages 7 and 8 are shown in **Table 1**. The average standard scores for receptive and expressive languages at both ages were all around 1 SD below the population mean, whilst average PIQ was above the population mean at both ages. All children had English as a first language. A small number of children, 27 (11.8%), had exposure to languages other than English at home. Household income data were available for 144 of the participants, with 9.7% coming from households with an annual income of less than £5500 (low income), 24.3% from households earning between £5600 and £10,500 (low-medium income), 19.4% from households earning between £10,600 and £15,500 (medium income), 22.9% from households with an annual income of between £15,600 and £21,000 (medium–high income), and 23.6% from families with an income of over £21,000 (high income).

MEASURES
Basic number skills
Number abilities were assessed by Test B of the Basic Number Skills subtest of the British Ability Scales (BAS) (Elliot, 1983). The test covers both pre-numerical (matching, classifying) and

Table 1 | Psycholinguistics profiles at ages 7 and 8.

	Age 7			Age 8		
	N	**Mean**	*SD*	*N*	**Mean**	*SD*
Receptive language raw scores[a]	229	10.2	3.4	229	12.4	3.2
Receptive language standard scores[a]	229	84.0	10.9	229	86.1	12.3
Expressive language raw scores[b]	219	21.8	7.7	221	25.6	8.4
Expressive language standard scores[b]	219	83.7	10.5	221	84.2	11.4
Receptive and expressive language composite[c]	219	32.1	10.2	221	38.0	10.5
PIQ raw scores[d]	221	19.1	4.6	227	21.8	5.4
PIQ standard scores[d]	221	106.0	14.9	214	108.8	15.5

[a] TROG.
[b] Bus Story Test.
[c] Sum of TROG and Bus Story Test raw scores.
[d] Raven's Coloured Progressive Matrices.

numerical aspects of number skills. Questions are asked covering the following areas: (1) matching and classifying, using qualitative attributes (e.g., matching spades to buckets based on their sizes); (2) matching and classifying by number (e.g., matching 2 ladybirds which have the same number of spots); (3) one to one correspondence (e.g., matching number of boxes in the picture to numerals); (4) comparison of sets (e.g., comparing picture of carts with more, less and same number of boxes); (5) ordinal aspects of number; (6) knowledge of number names and numerals; (7) counting a set of objects; (8) awareness of number patterns; (9) place value, including the ability to count in tens and knowledge of place value notation; (10) basic understanding of the four arithmetical operations.

For all tasks, children are presented with color picture cues and the instructions of the test are given verbally by the assessor, following the wordings given in the manual. Gestures are also used, e.g., by pointing and circling objects in the booklet. For example, for the question on matching and classifying using qualitative attributes, two pictures cues are used, one with two spades (one big and one small) and one with eight buckets (four big and four small). The child is asked to match spades to buckets based on their sizes. Indicating appropriately, the tester says "Here is a big spade and here is a little spade." Indicating with a circular motion, s/he then says "Here are the buckets that go with them. Can you show me all the buckets that go with the big spade?" For matching and classifying by number, again using two pictures as cues, one with four yellow ladybirds and one with 12 red ladybirds, the child is asked to match a red and a yellow ladybird which have the same number of spots. The instructions of the test include, "Here are 4 yellow ladybirds," then indicating to the appropriate ladybirds, the tester says "They may have a lot of spots or (indicating appropriately) very few spots or something in between." Pointing to the yellow ladybird with nine spots, the tester then asks the child to "Find me the red ladybirds that go with this one." Similarly, to assess basic understanding of the four arithmetical operations, one of the questions involved showing the child a picture of a box with seven buttons. The tester then says to the child "There used to be 12 buttons in this box. How many have been taken out?"

Testing discontinued after six successive items have been failed. The number of correct answers is summed to give a basic number skills raw score, which can then be converted into a standard score and be assigned an age-relevant percentile.

Receptive and expressive language

Receptive language at ages 7 and 8 was assessed using the Test for Reception of Grammar (TROG; Bishop, 1982). This is a test of oral comprehension of syntax in which children are shown four pictures while the examiner reads a sentence. The child is asked to pick the picture that illustrates the sentence. These items begin very simply and progress to more complex grammatical sentences (e.g., "the cat the cow chases is black"). Items are organized into blocks of four grammatically related sentences. The number of blocks passed is noted to give the TROG raw score, which can then be transformed into a standard score. Expressive language at both ages was assessed using the Bus Story Test (Renfrew, 1991), which is part of the Renfrew Language Scales. In this

assessment, the examiner tells the child a short story about a bus while the child looks through a book of pictures illustrating the story. The child must then retell the story as accurately as possible using the pictures as cues. Stories are audiotaped, transcribed and scored for the amount of correct information given. Two points are given for information central to the story, and one point for peripheral details, and these are summed to give the "total information score," which can subsequently be converted into a standard score. The receptive and expressive language raw scores were found to be highly correlated, $r = 0.62$, $p < 0.001$ at age 7, and $r = 0.61$, $p < 0.001$ at age 8. For the purpose of this study, therefore, a composite score representing both receptive and expressive language ability was derived by summing the TROG raw score and the Bus Story "total information score" measured at the corresponding age.

Performance IQ

Raven's Coloured Progressive Matrices (Raven, 1986) was used to assess participants' PIQ at both ages. The test is designed for educational and clinical assessment. It can be used satisfactorily with children who have language impairment and is commonly used to assess the PIQ of children with SLI. This test presents the child with a series of patterns from which a "piece" is missing, and the child is asked to choose from six alternative pieces the one which completes the pattern. The test is split into three sets of 12 patterns each and the number of correct answers is summed. This total score is then compared with age-relevant population norms.

PROCEDURE

Following informed written consent from families, children were visited at school and assessed individually in a quiet room or area with only the participant and a trained researcher present. The battery of psychometric tests was completed as part of the wider study which also included assessments for vocabulary, reading, articulation, and grammatical knowledge. The assessments were conducted in the same order for every child: (1) TROG, (2) BAS Number Skills, (3) BAS Naming Vocabulary (Elliot, 1983), (4) BAS Word Reading (Elliot, 1983), (5) Goldman-Fristoe Test of Articulation (Goldman and Fristoe, 1986), (6) Raven's Coloured Progressive Matrices, (7) Illinois Test of Psycholinguistic Ability: grammatic closure (Kirk et al., 1968), and (8) Renfrew Bus Story Test. In nearly all cases, testing was completed in 1 day at the child's pace and with normal school breaks. Because of the large number of measures used, the numbers of data points available may vary from measure to measure. Both raw scores and standard scores were examined in relation to children's basic number skills. All subsequent analyses used raw scores. All statistical analyses were conducted using Stata/SE 12.0 (StataCorp, 2011). All figures were also plotted using Stata.

RESULTS
BASIC NUMBER SKILLS
Preliminary analyses showed no main effects or interactions involving participant gender or household income; hence, data were pooled across these variables for the main analyses. Descriptive statistics on BAS number skills (raw scores and standard scores) as a function of age are presented in **Table 2**.

Table 2 | Basic number skills at ages 7 and 8.

	Age 7				Age 8			
	Mean	SD	Min	Max	Mean	SD	Min	Max
BAS number skills raw scores	21.0	6.3	2	34	25.8	6.2	0	34
BAS number skills standard scores	82.3	13.0	64	121	80.1	13.4	64	112
BAS number skills percentile[a]	18.6	21.4	<1	92	16.7	19.4	<1	78

N = 229.

[a] 25 (10.9%) and 48 (21.0%) of the 229 children at ages 7 and 8, respectively scored less than 1 percentile. These cases were coded as having a percentile of 0.5 when the mean percentiles were calculated.

Table 3 | Distribution of basic number skills standard scores at ages 7 and 8.

Range of standard scores	Age 7 (%)	Age 8 (%)
≤70 (2 SD below mean)	16.6	30.1
>70 and ≤77.5 (1.5–2 SD below mean)	27.5	18.8
>77.5 and ≤85 (1–1.5 SD below mean)	23.1	17.5
>85 and ≤100 (Between 1 SD below mean and mean)	23.1	25.3
>100 and <115 (Above mean but below 1 SD above mean)	8.3	8.3
≥115 (1 SD above mean)	1.3	0
Total	100	100

Table 4 | Linear regression analysis predicting basic number skills performance using concurrent variables at ages 7 and 8.

	b	SE	β	R^2	Adjusted R^2	Cohen's f^2
RAW SCORES AT AGE 7 (N = 212)						
Step 1				0.170	0.166	0.20
PIQ at age 7	0.54***	0.08	0.41			
Step 2				0.366	0.360	0.58
PIQ at age 7	0.24**	0.08	0.18			
Receptive and expressive language composite at age 7	0.30***	0.04	0.50			
TRANSFORMED RAW SCORES AT 8[a] (N = 220)						
Step 1				0.230	0.226	0.30
PIQ at age 8	−0.09***	0.01	−0.48			
Step 2				0.403	0.398	0.68
PIQ at age 8	−0.05***	0.01	−0.25			
Receptive and expressive language composite at age 8	−0.05***	0.01	−0.48			

**p < 0.01.

***p < 0.001.

[a] Reflect square root transformation of raw scores at age 8.

The raw scores rose by an average of 4.8 points between ages 7 and 8, $t_{(228)} = 14.9$, $p < 0.001$, Cohen's $d = 0.99$. The mean standard scores were more than 1 SD below the population mean at both ages. The standard scores declined on average 2.2 points, a significant drop, $t_{(228)} = −3.2$, $p = 0.001$, Cohen's $d = 0.2$. Thus, although there had been an improvement in the participants' basic number skills during the year, compared to normative data for TD peers they were on average performing more poorly at age 8 than at age 7. This drop in performance relative to the general population can also be seen in **Table 3**, which shows the distribution of the standard scores for the two ages. For example, while 16.6% of the children scored 2 SD below the population mean at age 7, this rose to 30.1% by age 8.

CONCURRENT PREDICTORS OF BASIC NUMBER SKILLS

Concurrent predictors of performance (using raw scores) at each age were examined first. Using linear regression analysis, covariates were added into the models in two steps: PIQ in the first step, followed by language skills. Given that the distribution of the basic number skills raw scores at age 8 was negatively skewed (skewness = −0.70, kurtosis = 3.31), reflect square root transformations were carried out and the transformed variables were used as the outcome in the linear regression at age 8.

The regression results, shown in **Table 4**, indicated that PIQ and language scores were significant predictors of concurrent basic number skills at both ages, with better abilities in these areas being associated with better basic number skills. The adjusted R^2 values showed that together they accounted for 36 and 40% of the variance in basic number skills performance at ages 7 and 8,

respectively. Comparisons of the standardized regression coefficients also suggested that the effect of language was greater than that of PIQ. After controlling for the effect of PIQ, language skills explained an additional 19 and 17% of the variance in the raw scores for ages 7 and 8, respectively. **Figures 1A,B** show the fitted (i.e., predicted) values of basic number skills raw scores vs. the language composite, for ages 7 and 8, respectively. At age 8, since the reflect square root transformed number skills raw scores were used as the outcome variable in the linear regression, for easier graphical interpretation, we have re-transformed the predicted values back to the original scale of the raw scores. The figures show that for both ages, the better the language, the higher the predicted values of basic number skills (controlling for PIQ).

FIGURE 1 | (A,B) Fitted values of basic number skills raw scores vs. receptive and expressive language composite—concurrent predictors of basic number skills at ages 7 and 8.
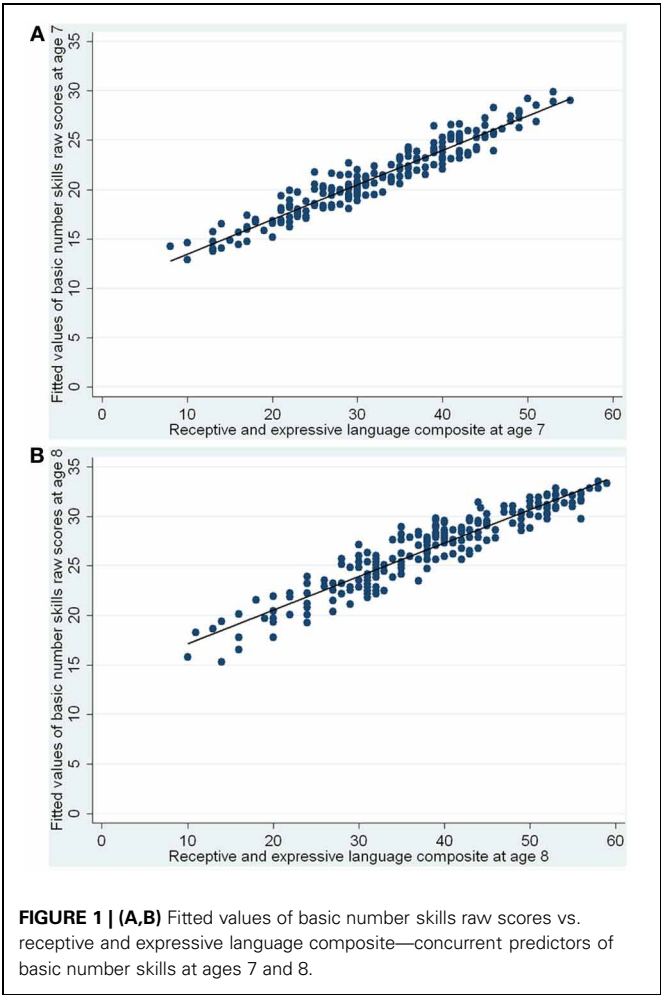
PREDICTING NUMBER SKILLS AT AGE 8 USING VARIABLES AT AGE 7

Using linear regression analyses and the approach of a conditional change model, we next investigated which variables at age 7 could predict basic number skills at 8. With the reflect square root transformed numbers skills raw scores at 8 as the outcome variable, PIQ raw scores were entered first, followed by basic number skills raw scores at 7, and lastly, the language composite. The results are shown in **Table 5**. Basic number skills at 7 and language skills at 7 are both significant predictors of number skills at 8, with better abilities in these areas at 7 associating with better basic number skills at 8. The standardized regression coefficients showed that the effect of earlier number skills was greater than that of earlier language skills. Having controlled for PIQ at 7, number skills at 7 accounted for 36% of the variance in basic number skills performance at 8, while language skills explained a further 4.6% of the remaining variance. The results are also illustrated in **Figure 2** which shows the fitted values of basic number skills raw scores at age 8 vs. basic number skills raw scores at age 7, and in **Figure 3**, which shows the fitted values of basic number skills raw scores at age 8 vs. language composite at age 7. As with **Figure 1B**, we have re-transformed the predicted values back to the original scale of the raw scores for easier interpretation. The figures again reveal that the higher the basic number skills raw scores at age 7 and the

Table 5 | Linear regression analysis predicting basic number skills performance at 8 using variables at age 7.

	b	SE	β	R^2	Adjusted R^2	Cohen's f^2
TRANSFORMED RAW SCORES AT 8[a]						
Step 1				0.161	0.157	0.19
PIQ at 7	−0.09***	0.01	−0.40			
Step 2				0.522	0.517	1.09
PIQ at 7	−0.03*	0.01	−0.13			
Basic number skills raw scores at 7	−0.11***	0.01	−0.66			
Step 3				0.570	0.564	1.32
PIQ at 7	−0.01	0.01	−0.06			
Basic number skills raw scores at 7	−0.09***	0.01	−0.53			
Receptive and expressive language composite at 7	−0.03***	0.01	−0.28			

$N = 212$.

[a] Reflect square root transformation of raw scores at age 8.

*$p < 0.05$.

***$p < 0.001$.

FIGURE 2 | Fitted values of basic number skills raw scores at age 8 vs. basic number skills raw scores at age 7.
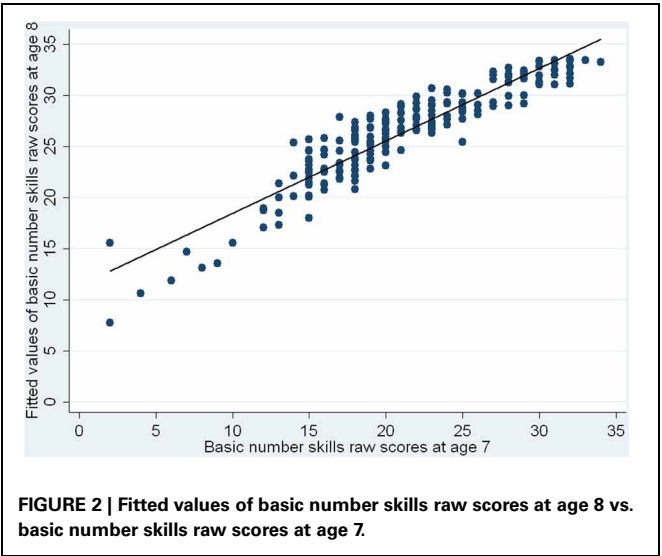
better the language abilities at age 7, the better the basic number skills at age 8.

PREDICTING THE DROP IN PERFORMANCE IN BASIC NUMBER SKILLS BETWEEN AGES 7 AND 8

Some improvements in most children's scores could be expected over the course of 1 year. However, some participants failed to show improvement, and some showed a decrease. Investigating why some children are not making progress, or appear to decline,

FIGURE 3 | Fitted values of basic number skills raw scores at age 8 vs. receptive and expressive language composite at age 7.
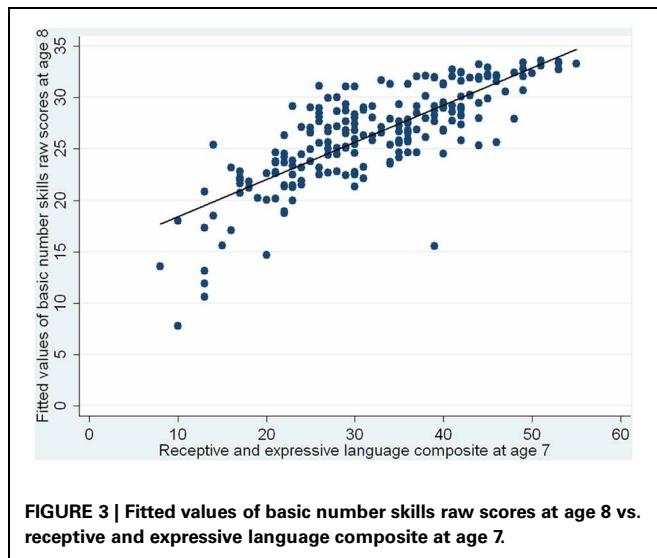
Table 6 | Participants showing a change or no change in the basic number skills raw and standard scores between ages 7 and 8.

	N	%	Mean change	SD
RAW SCORES				
No change in raw scores	12	5.2	0	—
Decrease in raw scores at 8	24	10.5	−3.5	2.4
Increase in raw scores at 8	193	84.3	6.1	4.0
All	229	100	4.8	4.9
STANDARD SCORES				
No change	31	13.5	0	—
Decrease in standard scores at 8	122	53.3	−9.8	6.5
Increase in standard scores at 8	76	33.2	9.0	6.6
All	229	100	−2.2	10.5

in a particular skills area contributes importantly to our understanding of which factors impede development. Hence, we examined the proportions of children who, from 7 to 8 years, showed no change, or who showed a decrease in scores, or who showed an increase in scores. These proportions were examined with respect to raw scores and standard scores. The results are presented in **Table 6**. Although 84% had seen an increase in their raw scores, only a third (33%) had seen a rise in their standard scores. In fact, the standard scores of over half of the children (53%) had fallen between the two ages. This again suggested that although the basic number skills for the vast majority of participants had progressed between ages 7 and 8, the rate of progress, on average, was not as great as that of their population peers.

Binary logistic regression analyses were carried out to examine which variables could predict a drop in the maths scores between ages 7 and 8. We focus on children who remain static or show a decline on the grounds that these are most likely to yield information on which factors impede development. These are also the children most in need of educational support. The outcome of the regression was coded as "0" (the reference category), representing no change or an increase in raw scores, or "1," a decrease

Table 7 | Binary logistic regression analysis predicting a drop in performance in basic number skills between ages 7 and 8.

	Odds ratio	SE	95% CI
STEP 1			
Difference in PIQ raw scores	0.88*	0.05	0.78–0.99
STEP 2			
Difference in PIQ raw scores	0.90	0.06	0.80–1.02
Difference in language composites	0.90**	0.03	0.84–0.97

$N = 208$; $*p < 0.05$; $**p < 0.01$.

in raw scores. A variable representing the difference in PIQ raw scores between ages 7 and 8 was first entered into the model, followed by the language composite raw score differences between 7 and 8 years. The results for predicting a drop in performance in basic number abilities are presented in **Table 7**. The difference in language skills between 7 and 8 years was the only significant predictor in the final model. With an odds ratio of less than 1, results of the final model suggested that the larger the change in language skills between ages 7 and 8, the smaller the odds of a drop in basic number abilities. That is, the greater the improvement in language ability, the lesser the likelihood of a drop in performance in basic number skills from 7 to 8 years: less language improvement, more likelihood of a drop in language skills. With an odds ratio of 0.90, one unit decrease in the difference in language composites was associated with a 11% increase in the odds of having a drop in number skills at age 8. The analogous analysis was performed on standard scores and the pattern of results was essentially unchanged.

DISCUSSION

The purpose of this longitudinal study was to examine the extent to which, in a sample of children with SLI, the severity of their condition predicts progress in number skills development during the early primary school years. Adding to earlier evidence of a disadvantage to children with SLI in this domain (Fazio, 1999; Arvedson, 2002; Cowan et al., 2005; Donlan et al., 2007; Kleemans et al., 2011, 2012), we obtained clear findings that this large sample obtained mean standard scores more than 1 SD below the population mean at age 7 and again at age 8. The standard scores of over half of the children (53%) had fallen between the two ages. Overall, then, children with SLI fell well below TD norms at age 7, and the gap worsened by age 8.

Both PIQ and language scores were significant predictors of concurrent number skills at both ages, with better abilities in these areas being associated with better basic number skills. Together, these variables accounted for 36 and 40% of the variance in basic number skills performance at ages 7 and 8, respectively. After controlling for the effect of PIQ, language skills explained an additional 19 and 17% of the variance in the raw scores for ages 7 and 8, respectively. Cowan et al. (2005) found a similar pattern in their analysis of data pooled across TD and SLI participants; the present findings confirm that the relationship holds within SLI. Thus, severity of language impairment is clearly associated with number skills ability.

The next question of interest was whether language ability at age 7 contributed to the prediction of number scores at age 8. We regressed number scores at age 8 on PIQ at age 7, maths score at age 7, and language ability at age 7. This analysis revealed that, after taking account of the other predictors, i.e., PIQ and maths scores, language abilities explained an additional 4.6% of variance in the later number scores. This indicates that, as well as impacting at age 7, severity of language impairment also influences continuing progress in number skills over the following year.

We examined next whether the likelihood of decline in number scores could be predicted from changes in language ability, controlling for changes in PIQ. The change in language skills between 7 and 8 years emerged as the only significant predictor in this analysis. The results were robust and indicated that less improvement in language ability was associated with a greater odds of a drop in performance in basic number skills from 7 to 8 years.

With respect to the contrasting theoretical positions we outlined above, the evidence does not support a strong version of the "separate origins" hypothesis, holding that number and language abilities develop independently. This would predict that progress in number performance should not be affected by the severity of language impairment and the present results support the opposite conclusion. However, the present study, focusing on middle childhood, is moot on the question of early origins and, because all of the number tasks we administered required at least some degree of verbal processing, we do not have data to allow us to compare performance on non-verbal number tasks. Previous research has addressed this issue and found mixed results, with some indicating that children with SLI can perform at around the level of typical peers on non-verbal number tests (Arvedson, 2002; Donlan et al., 2007; Kleemans et al., 2011, 2012), and others suggesting more pervasive deficits (Cowan et al., 2005).

Our results are more readily compatible with theoretical stances holding that the development of number skills is influenced by the child's linguistic abilities. Consistent with this position, our findings that children with SLI are at a disadvantage and show an increased gap from peer norms over the course of a year suggest that language is an impediment to their early engagement with number and a continuing liability as children with SLI face increasingly complex number work in primary education. Further research is clearly warranted to determine if non-verbal number skills progress at a different pace to verbally-mediated number skills in children with SLI, but it is important to bear in mind that, in practice, school work in number and mathematics tends to be immersed in language, both domain-specific and general (Durkin and Shire, 1991; Ginsburg, 2009).

Evidence from children with SLI provides valuable information to help our understanding of both atypical and typical development. The present findings confirm that the child's ability to handle the medium of communication bears on his or her progress in a crucial area of learning, number skills development. The more severe the language impairment, the more difficult the task of mastering number skills. Conversely, we can infer that unimpaired language development in the typical child supports number and mathematical development. This is less readily apparent in typical development because the language ability is assumed and transparent, but recall that Cowan et al. (2005) found that language comprehension was the best predictor of variation on number tasks in a combined sample of TD and SLI participants. Of course, many other endogenous and environmental factors also bear on the complex processes of learning about number, but language is a fundamental component.

It should be acknowledged that the origins of language impairment itself have not been addressed here. It could be argued that both language and number skills development are dependent on similar underlying neural and cognitive processes, such as working memory. This in turn raises questions about the nature of working memory resources (e.g., verbal vs. non-verbal) that might be drawn upon in number tasks (see Nys et al., 2013, for an interesting discussion). While these are broader issues than our present concerns, we reiterate that evidence from children with SLI, who show language difficulties in the context of cognition within the normal range, makes an important contribution to the larger inquiry into the underpinnings of number development.

The present study did not include an age-matched TD sample. Several previous studies have done so and the findings are clear. An advantage of such designs is that the participants are tested contemporaneously, by the same investigators under similar conditions; a possible disadvantage is that the experimenters are not usually blind to participant diagnostic status, raising the possibility of demand characteristics contributing to results. By using a standardized test of number abilities we were able to draw instead on normative data collected from a large representative sample, independently of the present investigators. The fact that similar results are obtained via each comparison strategy supports confidence in the overall pattern of results.

There are obvious but important practical implications that follow from our findings. We found that not only did children with SLI tend to underperform in number skills, relative to standard performance, some remained static over a full school year and a substantial proportion declined. There is the risk that the problems children with SLI face in early mathematics education are compounded over time. The tasks themselves become more advanced and, as others have pointed out (Arvedson, 2002; Cowan et al., 2005), children's difficulties with the verbal demands may be less apparent to educators and therapeutic intervention is likely to be concentrated on linguistic skills. Future research is needed to investigate whether the worsening gap that we have demonstrated here continues to grow over the remainder of primary schooling and beyond, and to assess the efficacy of interventions designed to meet these children's language needs in the particular context of mathematics.

In sum, the present study, with one of the largest SLI samples to be tested on number development in childhood, reinforces earlier findings that children with this disorder are at a disadvantage in this domain. Important novel contributions are that not only do children with SLI fall below peer norms on number tasks

but that the gap worsens from age 7 to 8, that language ability contributes substantially to the concurrent prediction of number scores, that language ability at age 7 contributes to the prediction of number scores a year later, and that changes in language ability over the year are associated with whether or not children with SLI manifest a drop in performance in basic number skills at age 8. Language ability underpins progress in number understanding and education.

ACKNOWLEDGMENTS

REFERENCES

Ansari, D., and Karmiloff-Smith, A. (2002). Atypical trajectories of number development: a neuroconstructivist perspective. *Trends Cogn. Sci.* 6, 511–516. doi: 10.1016/S1364-6613(02)02040-5

Arvedson, P. J. (2002). Young children with specific language impairment and their numerical cognition. *J. Speech Lang. Hear. Res.* 45, 970–982. doi: 10.1044/1092-4388(2002/079)

Bishop, D. V. M. (1982). *Test for Reception of Grammar*. Manchester: Author.

Bishop, D. V. M. (1997). *Uncommon Understanding: Development and Disorders of Language Comprehension in Children*. Hove: Psychology Press.

Bishop, D. V. M., and Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *J. Child Psychol. Psychiatry* 31, 1027–1050. doi: 10.1111/j.1469-7610.1990.tb00844.x

Bloom, P., and Wynn, K. (1997). Linguistic cues in the acquisition of number words. *J. Child Lang.* 24, 511–533. doi: 10.1017/S0305000997003188

Carey, S. (2004). Bootstrapping and the origin of concepts. *Daedalus* 133, 59–68. doi: 10.1162/001152604772746701

Conti-Ramsden, G., and Botting, N. (1999). Characteristics of children attending language units in England: a national study of 7-year-olds. *Int. J. Lang. Commun. Disord.* 34, 359–366. doi: 10.1080/136828299247333

Conti-Ramsden, G., Crutchley, A., and Botting, N. (1997). The extent to which psychometric tests differentiate subgroups of children with specific language impairment. *J. Speech Lang. Hear. Res.* 40, 765–777.

Conti-Ramsden, G., and Durkin, K. (2011). "Specific language impairment," in *Child Psychology and Psychiatry: Frameworks for Practice*, eds D. Skuse, H. Bruce, L. Dowdney, and D. Mrazek (Chichester: John Wiley and Sons), 180–186. doi: 10.1002/9781119993971.ch29

Conti-Ramsden, G., Durkin, K., Simkin, Z., and Knox, E. (2009). Specific language impairment and school outcomes I: identifying and explaining variability at the end of compulsory education. *Int. J. Lang. Commun. Disord.* 44, 15–35. doi: 10.1080/13682820801921601

Conti-Ramsden, G., St. Clair, M. C., Pickles, A., and Durkin, K. (2012). Developmental trajectories of verbal and nonverbal skills in individuals with a history of specific language impairment: from childhood to adolescence. *J. Speech Lang. Hear. Res.* 55, 1716–1735. doi: 10.1044/1092-4388(2012/10-0182)

Cordes, S., and Brannon, E. M. (2009). Crossing the divide: infants discriminate small from large numerosities. *Dev. Psychol.* 45, 1583–1594. doi: 10.1037/a0015666

Cowan, R., Donlan, C., Newton, E. J., and Lloyd, D. (2005). Number skills and knowledge in children with specific language impairment. *J. Educ. Psychol.* 97, 732–744. doi: 10.1037/0022-0663.97.4.732

Donlan, C., Cowan, R., Newton, E. J., and Lloyd, D. (2007). The role of language in mathematical development: evidence from children with specific language impairments. *Cognition* 103, 23–33. doi: 10.1016/j.cognition.2006.02.007

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Dev. Psychol.* 43, 1428–1446. doi: 10.1037/0012-1649.43.6.1428

Durkin, K. (1993). "The representation of number," in *Systems of Representation in Children. Development and Use*, eds C. Pratt and A. Garton (Chichester: Wiley), 133–166.

Durkin, K., and Shire, B. (1987). Primary school children's interpretations of lexical ambiguity in mathematical descriptions. *J. Res. Read.* 32, 43–50.

Durkin, K., and Shire, B. (eds.). (1991). *Language in Mathematical Education: Research and Practice*. Milton Keynes: Open University Press.

Durkin, K., Shire, B., Riem, R., Crowther, R. D., and Rutter, D. R. (1986). The social and linguistic context of early number word use. *Br. J. Dev. Psychol.* 4, 269–288. doi: 10.1111/j.2044-835X.1986.tb01018.x

Durkin, K., Simkin, Z., Knox, E., and Conti-Ramsden, G. (2009). Specific language impairment and school outcomes II: educational context, student satisfaction, and post-compulsory progress. *Int. J. Lang. Commun. Disord.* 15, 36–55. doi: 10.1080/13682820801921510

Ellerton, N. F., and Clarkson, P. C. (1996). "Language factors in mathematics teaching and learning," in *International Handbook of Mathematics Education (Part 2)*, eds A. J. Bishop, M. A. Clements, C. Keitel, J. Kilpatrick, and C. Laborde (Dordrecht: Kluwer Academic Publishers), 987–1033.

Elliot, C. D. (1983). *British Ability Scales*. Windsor: NFER-Nelson.

Fazio, B. B. (1994). The counting abilities of children with specific language impairment: a comparison of oral and gestural tasks. *J. Speech Lang. Hear. Res.* 37, 358–368.

Fazio, B. B. (1996). Mathematical abilities of children with specific language impairment: a 2-year follow-up. *J. Speech Lang. Hear. Res.* 39, 1–10.

Fazio, B. B. (1999). Arithmetic calculation, short-term memory, and language performance in children with specific language impairment: a 5-year follow-up. *J. Speech Lang. Hear. Res.* 42, 420–431.

Fuson, K. C. (1988). *Children's Counting and Concepts of Number*. New York, NY: Springer Verlag. doi: 10.1007/978-1-4612-3754-9

Gelman, R., and Butterworth, B. (2005). Number and language: how are they related? *Trends Cogn. Sci.* 9, 6–10. doi: 10.1016/j.tics.2004.11.004

Ginsburg, H. P. (2009). "Early mathematics education and how to do it," in *Handbook of Child Development and Early Education: Research to Practice*, eds O. A. Barbarin and B. H. Wasik. (New York: Guilford Press), 423–428.

Goldman, R., and Fristoe, M. (1986). *Goldman-Fristoe Test of Articulation*. Circle Pines, MN: American Guidance Association.

Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). Language effects on children's non-verbal number line estimations. *J. Cross Cult. Psychol.* 42, 598–613. doi: 10.1177/0022022111406026

Izard, V., Sann, C., Spelke, E. S., and Streri, A. (2009). Newborn infants perceive abstract numbers. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10382–10385. doi: 10.1073/pnas.0812142106

Kirk, S. A., McCarthy, J. J., and Kirk, W. D. (1968). *Illinois Test of Psycholinguistic Abilities*, Urbana, IL: University of Illinois.

Kleemans, T., Segers, E., and Verhoeven, L. (2011). Precursors to numeracy in kindergartners with specific language impairment. *Res. Dev. Disabil.* 32, 2901–2908. doi: 10.1016/j.ridd.2011.05.013

Kleemans, T., Segers, E., and Verhoeven, L. (2012). Naming speed as a clinical marker in predicting basic calculation skills in children with specific language impairment. *Res. Dev. Disabil.* 33, 882–889. doi: 10.1016/j.ridd.2011.12.007

Levine, S. C., Suriyakham, L. W., Rowe, M. L., Huttenlocher, J., and Gunderson, E. A. (2010). What counts in the development of young children's number knowledge? *Dev. Psychol.* 46, 1309–1319. doi: 10.1037/a0019671

Linnell, M., and Fluck, M. (2001). The effect of maternal support for counting and cardinal understanding in pre-school children. *Soc. Dev.* 10, 202–220. doi: 10.1111/1467-9507.00159

Mix, K. S. (2009). How Spencer made number: first uses of the number words. *J. Exp. Child Psychol.* 102, 427–444. doi: 10.1016/j.jecp.2008.11.003

Mix, K. S., Sandhofer, C. M., and Baroody, A. J. (2005). "Number

words and number concepts: the interplay of verbal and nonverbal processes in early quantitative development," in *Advances in Child Development and Behavior,* Vol. 33, ed R. Kail (New York, NY: Academic Press).

Mix, K. S., Sandhofer, C. M., Moore, J. A., and Russell, C. (2012). Acquisition of the cardinal word principle: the role of input. *Early Child. Res. Q.* 27, 274–283. doi: 10.1016/j.ecresq.2011.10.003

Musolino, J. (2004). The semantics and acquisition of number words: integrating linguistic and developmental perspectives. *Cognition* 93, 1–41. doi: 10.1016/j.cognition.2003.10.002

Negen, J., and Sarnecka, B. W. (2012). Number-concept acquisition and general vocabulary development. *Child Dev.* 83, 2019–2027. doi: 10.1111/j.1467-8624.2012.01815.x

Nelson, K. E., Welsh, J. A., Vance Trup, E. M., and Greenberg, M. T. (2011). Language delays of impoverished preschool children in relation to early academic and emotion recognition skills. *First Lang.* 31, 164–194. doi: 10.1177/0142723710391887

Nieder, A., and Dehaene, S. (2009). Representation of number in the brain. *Annu. Rev. Neurosci.* 32, 185–208. doi: 10.1146/annurev.neuro.051508.135550

Núñez, R. (2011). No innate number line in the human brain. *J. Cross Cult. Psychol.* 45, 651–668.

Nys, J., Content, A., and Leybaert, J. (2013). Impact of language abilities on exact and approximate number skills development: evidence from children with specific language impairment. *J. Speech Lang. Hear. Res.* 56, 956–970. doi: 10.1044/1092-4388(2012/10-0229)

O'Neill, D. K., Pearce, M. J., and Pick, J. L. (2004). Preschool children's narratives and performance on the Peabody Individualized Achievement Test—Revised: evidence of a relation between early narrative and later mathematical ability. *First Lang.* 24, 149–183. doi: 10.1177/0142723704043529

Palmer, A., and Baroody, A. J. (2011). Blake's development of the number words "one," "two," and "three." *Cogn. Instr.* 29, 265–296. doi: 10.1080/07370008.2011.583370

Pixner, S., Moeller, K., Hermanova, V., Nuerk, H. C., and Kaufmann, L. (2011). Whorf reloaded: language effects on non-verbal number processing in 1st grade—a trilingual study. *J. Exp. Child Psychol.* 108, 371–382. doi: 10.1016/j.jecp.2010.09.002

Raven, J. C. (1986). *Coloured Progressive Matrices*. London: H. K. Lewis.

Reikerås, E., Løge, I. K., and Knivsberg, A.-M. (2012). The mathematical competencies of toddlers expressed in their play and daily life activities in Norwegian kindergartens. *Int. J. Early Child.* 44, 91–114. doi: 10.1007/s13158-011-0050-x

Renfrew, C. (1991). *The Bus Story: A Test of Continuous Speech*. Oxford: author.

Roberts, S. H., and Gathercole, V. C. M. (2006). Categorizing collections of objects: linguistic and cognitive factors influencing Welsh and English speakers' judgments. *First Lang.* 26, 161–185. doi: 10.1177/0142723706060745

Salehuddin, K., and Winskel, H. (2009). An investigation into Malay numeral classifier acquisition through an elicited production task. *First Lang.* 29, 291–313. doi: 10.1177/0142723709103187

Seron, X., and Fayol, M. (1994). Number transcoding in children: a functional analysis. *Br. J. Dev. Psychol.* 12, 281–300. doi: 10.1111/j.2044-835X.1994.tb00635.x

Sophian, C. (2007). *The Origins of Mathematical Knowledge in Childhood*. New York, NY: Lawrence Erlbaum Associates.

Spelke, E., and Tsivkin, S. (2001). "Initial knowledge and conceptual change: space and number," in *Language Acquisition and Conceptual Development*, eds M. Bowerman and S. C. Levinson (Cambridge: Cambridge University Press), 70–100. doi: 10.1017/CBO9780511620669.005

StataCorp (2011). *Stata Statistical Software: Release 12.0* [Computer Software] College Station, TX: Stata Corp LP.

St. Clair, M. C., Durkin, K., Conti-Ramsden, G., and Pickles, A. (2010). Growth of reading skills in children with a history of specific language impairment (SLI): the role of autistic symptomatology and language related abilities. *Br. J. Dev. Psychol.* 28, 109–131. doi: 10.1348/026151009X480158

Strauss, M. S., and Curtis, L. E. (1981). Infant perception of numerosity. *Child Dev.* 52, 1146–1152. doi: 10.2307/1129500

Tare, M., Shatz, M., and Gilbertson, L. (2008). Maternal uses of non-object terms in child-directed speech: color, number, and time. *First Lang.* 28, 87–100. doi: 10.1177/0142723707085316

Tomblin, J. B., Records, N. L., Buckwalter, P. R., Zhang, X., Smith, E., and O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *J. Speech Lang. Hear. Res.* 40, 1245–1260.

Wiese, H. (2003). *Numbers, Language, and the Human Mind*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486562

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

Zuber, J., Pixner, S., Moeller, K., and 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. *J. Exp. Child Psychol.* 102, 60–77. doi: 10.1016/j.jecp.2008.04.003

The role of fingers in number processing in young children

Anne Lafay[1], Catherine Thevenot[2]*, Caroline Castel[2] and Michel Fayol[3]

[1] Centre de Recherche de l'Institut Universitaire en Santé Mentale de Québec, Université Laval, Québec, QC, Canada
[2] Faculté de Psychologie et des Sciences de l'éducation, Department of Psychology, Université de Genève, Genève, Suisse
[3] Centre National de la Recherche Scientifique, LAPSCO, Université Blaise Pascal, Clermont-Ferrand, France

The aim of the present study was to investigate the relationship between finger counting and numerical processing in 4–7-year-old children. Children were assessed on a variety of numerical tasks and we examined the correlations between their rates of success and their frequency of finger use in a counting task. We showed that children's performance on finger pattern comparison and identification tasks did not correlate with the frequency of finger use. However, this last variable correlated with the percentages of correct responses in an enumeration task (i.e., Give-N task), even when the age of children was entered as a covariate in the analysis. Despite this correlation, we showed that some children who never used their fingers in the counting task were able to perform optimally in the enumeration task. Overall, our results support the conclusion that finger counting is useful but not necessary to develop accurate symbolic numerical skills. Moreover, our results suggest that the use of fingers in a counting task is related to the ability of children in a dynamic enumeration task but not to static tasks involving recognition or comparison of finger patterns. Therefore, it could be that the link between fingers and numbers remain circumscribed to counting tasks and do not extent to static finger montring situations.

Keywords: numerical cognition, numerosity, preschoolers, non-symbolic representations, kindergartens

INTRODUCTION

Numerical symbols appear in a large variety of contexts such as price tags, shopping bills, phone numbers, street addresses, or arithmetic and mathematical problems. It is therefore important for researchers and practitioners to understand how numerical capacities develop and potentially dysfunction in children. Then, early identification of numerical difficulties has become a challenging and promising domain of research in order to construct and apply appropriate reeducation programs.

An early numerical ability that has received increasing attention from researchers in recent years is finger counting. As reminded by Dantzig (1962), Butterworth (1999) noted that "Whenever a counting technique, worthy of the name, exists at all, finger-counting has been found either to precede it or accompany it." Then, finger counting constitutes an external aid to represent numbers, helps keeping track of number words in counting and sustains the comprehension of the 10-base numerical system as well as the realization of basic arithmetic operations. Due to the amount of activities based on finger counting, it is logically considered to play an important role in numerical capacities.

The relationship between fingers and numerical representations has been established in several studies involving children with manual difficulties. Arp and Fagard (2001) showed that counting difficulties in children with cerebral palsy depend on the severity of their visual-manual coordination deficits. It has also been shown that children with dyspraxia present a delay in mathematical acquisition, which could be due to their difficulties in pointing at objects and would, in turn, prevent them from counting collection appropriately (Lecointre et al., 2005). If, in those studies, the difficulties encountered by children result from

visuo-manual coordination impairments, more recent studies isolate the role of manual deficits by assessing numerical abilities in children without coordination problems. Indeed, Thevenot and Fluss (2012) showed that children with congenital hemiplegia, who present difficulties in using one of their hands, also exhibit difficulties in symbolic numerical tasks. The relationship between numbers and fingers has also been well documented outside the field of neuropsychology. Fayol et al. (1998) showed that finger recognition performance in 5–6-year-old children is a better predictor of arithmetical performance one year later than more classical tests of intelligence such as Goodenough's draw-a-man test. This held true even three years later (Marinthe et al., 2001). Moreover, Costa et al. (2011) showed that finger gnosis performance is lower in children with mathematical difficulties than in children without difficulties. These results are coherent with those of Noël (2005) who revealed a positive correlation between finger gnosis and numerical performance in children at the beginning of Grade 1 (i.e., 6–7-year-old children). Gracia-Bafalluy and Noël (2008) even suggested that training finger gnosis may generalize to untrained numerical performance (but see Fischer, 2010).

A simple explanation of the relationship between numbers and fingers is given by the proximity of the brain areas devoted to their mental processing. This proximity has been suspected for the first time following Gerstmann's description (1940) of a series of patients who presented strange concomitant symptoms of agraphia, spatiotemporal confusion, finger agnosis, and acalculia. This syndrome, now known as the Gerstmann syndrome, has also been identified in children (Kinsbourne et al., 1963). Later on, neuro-imaging studies confirmed Gerstmann's intuition

showing that the parietal lobe and the left precentral gyrus are involved both in numerical processing and finger gnosia (Di Luca et al., 2006; Sandrini and Rusconi, 2009). Therefore, a lesion of the left parietal lobe can affect both representations of fingers and representations of numerosities. This was also nicely confirmed by a rTMS study showing that an angular gyrus stimulation generates disruptions in tasks involving finger and numerical representations (Rusconi et al., 2005).

Going further than a mere explanation of the relationship between numbers and fingers by the proximity of the brain areas devoted to these activities, a functionalist interpretation postulates that finger counting constitutes the basis for future numerical abilities (see Fayol and Seron, 2005 for a review). Within this interpretation, the link between finger gnosia and numerical abilities is well explained by Reeve and Humberstone (2011) who showed that finger gnosia abilities change in the early school years and that these changes are associated with the ability to use fingers to aid computation. Furthermore and in accordance with theories of embodied cognition (Barsalou, 1999), Domahs et al. (2008) observed children's errors in addition and subtraction problems and showed that split-five errors were over-represented and above chance level. The authors concluded that mental representations of numbers that inherit sub-base five properties are built up and internalized during childhood. Interestingly, Domahs et al. (2010) also showed the influence of the sub-base five on numerical representations in adults. More generally, within the "manumerical" hypothesis (Fischer and Brugger, 2011), finger-based representations of numbers are seen as the result of an integration of multi-modal input during early finger counting and finger calculation in childhood and the later offline simulation of the according motor programs (Moeller et al., 2012). Moeller et al. (2012) even suggest that finger-based representations of numbers are activated automatically whenever a number is encountered. Also in line with the functionalist interpretation, Andres et al. (2007) suggested that finger use in numerical tasks could constitute a transition between the non-symbolic and symbolic systems and would therefore determine later performance in representation and manipulation of numbers in a pure symbolic format. The authors confirmed their proposition by showing corticospinal excitability of the muscles of the hand in silent numerical tasks in adults. This was interpreted as evidence for childhood reminiscence of finger use for the representation of number words. As noted by Andres et al. (2008), fingers may be the "missing tool" between non-symbolic and symbolic numerosities involved in arithmetic.

Therefore, the fact that finger use in counting shapes numerical mental representations seems well supported in the literature. However, whether or not finger counting is a necessary tool for the development of these representations, and consequently numerical abilities, is still open for debate (Plaisier and Smeets, 2011). As noted recently by Crollen et al. (2011b), the functional hypothesis would lead to the prediction that, during the first developmental stages, children should be more accurate to represent numerosities with their fingers than with number words. As a matter of fact, no such data is available so far in the literature. On the contrary, Nicoladis et al. (2010) showed that 2–5-year-old children performed equally bad when presented

either with hand shape or number words and asked to put the corresponding number of toys in a box. Moreover, 4 and 5-year old children perform actually better with words than hand shape. Another interesting result that questions the role of fingers in shaping numerical abilities has been recently reported by Crollen et al. (2011a) who demonstrated that blind children use finger-counting strategies less often than sighted children. Still, blind and sighted children achieve similar level of performance in enumeration tasks. Crollen et al. concluded that fingers are a useful rather than necessary tool for the development of counting abilities.

These series of results question the necessity of finger use in the development of numerical abilities and suggest that further investigation is needed in order to determine the precise role of fingers in number processing. This is the aim of the present study. If finger use is indeed a useful step that constitutes a transient developmental stage between non-symbolic and symbolic numerical abilities, children who use their fingers less frequently should be those children with the poorer numerical performance. However and furthermore, if finger use is a necessary step in the numerical developmental course, children who do not use their fingers should not be able to succeed in symbolic numerical tasks. In order to verify this assumption, we assessed 4–7-year-old-children's performance on their spontaneous use of fingers in a counting task, on finger numerical pattern recognition and on an enumeration task. Seldom reliance on fingers for the counting task should be associated with poorer performance in finger pattern comparison and identification, and, in turn poorer performance in a "Give-me N" task (i.e., enumeration task).

The task that we developed in order to determine whether children use their fingers in a counting task is original and allows us to determine precisely whether finger use in numerical tasks is a strategy that belongs to children's repertoire. Children had to determine the total number of pictures in a collection presented in front of them on a table. They were asked to name the picture one by one, and, just after, they had to give the cardinal of the collection. Because, in this task, the phonological loop is blocked by picture naming (Baddeley, 1986), the best strategy in absence of other external aids is to keep track of the number of pictures on fingers. Of course, it was never mentioned to children that they had to (or even could) use their fingers to perform the task. Therefore, children who implemented the finger strategy did it spontaneously, without any constraint or insight from the experimenter. We think that this task is a better way to assess finger use in numerical activities than a mere observation of children's behavior during calculations. Indeed, when children do not use their fingers to solve arithmetic problems, it is impossible to determine whether they do not need them any longer or whether they never have resorted to them. On the contrary, in the picture counting task, fingers are still required to succeed in the task. Then, children who do not use them are necessarily children for whom this strategy is not available.

METHODS
PARTICIPANTS
Sixty normally developing children took part in this experiment. Twenty of them were preschoolers aged between 4 and 5

years ($M = 4.7$, $SD = 0.31$; 9 females; 18 right-handed). Twenty of them were kindergarten children aged between 5 and 6 years ($M = 5.6$, $SD = 0.29$; 9 females; 17 right-handed). The 20 remaining children were in Grade 1 and were aged between 6 and 7 years ($M = 6.7$, $SD = 0.29$; 11 females; 19 right-handed). Children did not present any developmental disorders or disabilities.

MATERIALS AND PROCEDURE
Spontaneous use of fingers in counting
As already mentioned above, children had to determine the total number of pictures in a collection presented in front of them on a table. They were asked to name the picture one by one, and, just after, they had to give the cardinal of the collection. We selected twenty pictures that lead to 100% of correct recognition in 4-year-old children (BD2I, Cannard et al., 2006). Children were presented with three small collections from 1 to 5 pictures, three medium collections from 6 to 10 pictures and three large collections from 11 to 15 pictures. For each child, the specific numerosities that were selected within each of the collection size were kept for the Give-*N* task (e.g., 2, 3, 5 for small; 6, 7, 10 for medium and 11, 14, 15 for large collections). Whether or not children used their fingers during the task, and whether or not they succeeded in giving the correct number of pictures presented constituted our two dependent variables of interest. The use of fingers was coded by the experimenter instantly during testing.

Finger numerical pattern recognition
Finger numerical pattern recognition was assessed with a comparison and an identification task. The materials we used for both tasks was adapted from Noël (2005; see also Gracia-Bafalluy and Noël, 2008).

Comparison of finger numerical pattern.
Sixteen different pictures representing one hand with one to four raised fingers were used in this task. Half of them represented right hands, while the other half represented left hands. More importantly, half of the pictures corresponded to canonical numerical finger patterns (e.g., culturally, raising the thumb, the index, and the middle finger represents 3 in France) and the other half to non-canonical patterns.

Twenty-four trials were constructed using those pictures. A trial corresponded to two pictures presented on screen at the same time. A third of the trials were composed of two pictures representing canonical patterns, another third of two pictures representing non-canonical patterns and a last third mixing canonical and non-canonical patterns (see **Table A1**). By pressing a key, children had to decide as quickly as possible if the two pictures showed the same number of raised fingers. Half of the trials required a "Yes" response and the other half a "No" response. Each trial was preceded by a fixation cross presented for 500 ms and the picture was displayed on screen until the answer was given. Four warm-up trials were presented before the experimental phase. Accuracy and reaction times were recorded by the computer.

Identification of finger numerical pattern.
In addition to the 16 pictures used in the previous task, 12 pictures representing two hands with six to nine raised fingers were added to the materials. The four canonical patterns corresponding to 6, 7, 8, and 9 were presented and, in order to increase the number of trials, two different non-canonical patterns were constructed for each numerosity (see **Table A2**). This resulted in twenty-eight experimental trials (i.e., 16 pictures representing one hand and 12 pictures representing two hands), which were preceded by three warm-up trials.

Children were asked to determine as quickly as possible the number of raised fingers on the pictures and the experimenter pressed a key as soon as participants uttered their answer. Reactions times were recorded by the computer, whereas errors were written down by the experimenter.

Enumeration task (give-N task)
Children had to give 1–15 tokens to the experimenter. The exact same numerosities than in the "Spontaneous use of fingers in counting" task were used here. As already explained, there were three trials per numerosity size (i.e., small numerosity: 1–5 tokens; medium numerosity: 6–10 tokens and large numerosity: 11–15 tokens). Children succeeded the task when they precisely took the number of tokens asked by the experimenter from a stack of tokens displayed in front of them.

RESULTS
In order to have a clear picture of children's behavior in the different tasks, an ANOVA will be performed for each of them. Moreover and in order to test our predictions, a correlation analysis between the frequency of spontaneous use of fingers and the other tasks will be reported for each of the tasks. Finally, in order to verify whether numerical abilities are really related to the use of fingers rather than to natural development, additional analyses will be conducted with the age of children as a covariate for each of the significant correlations.

SPONTANEOUS USE OF FINGERS IN COUNTING
Percentages of spontaneous finger use
Overall, 22 children out of 60 (36%) used their fingers at least once in the picture counting task. The number of children using their fingers increased as a function of age with only 1 preschooler (5%), 3 kindergartens (15%), and 18 children out of 20 in Grade 1 (90%) using them. For the sake of precision, the following ANOVA was carried out on a trial by trial basis and not on the rough percentages of children who used their fingers.

Because our dependent variable was binary, we applied an arcsin transformation to our data before carrying out the 3 (School level: Preschoolers, Kindergartens, and Grade 1) × 3 (Numerosities: Small, Medium, and Large) ANOVA on the transformed data (**Table 1**). Fingers were used in 29% of the trials and, in accordance with the previous results, the percentage of trials wherein fingers were used varied as a function of school level, $F_{(2, 57)} = 30.21$, $\eta_p^2 = 0.51$, $p < 0.001$. Children in Grade 1 used their fingers in 70% of the trials, whereas kindergartens and preschoolers used their fingers in only 12 and 5% of the trials, respectively. The percentages of trials wherein fingers were used also increased as a function of numerosities, $F_{(2, 114)} = 18.02$,

$\eta_p^2 = 0.24$, $p < 0.001$ (17, 34, and 35% for small, medium, and large numerosities, respectively). Moreover, there was an interaction between the two factors, $F_{(4, 114)} = 10.11$, $\eta_p^2 = 0.26$, $p < 0.001$, showing that the effect of school level increased with numerosities.

Percentages of correct responses in the counting task
As in the previous analysis, an arcsin transformation was applied to the percentages of correct responses in the counting task and an ANOVA with the same design as before was carried out on the transformed data (**Table 2**). The percentages of correct responses increased as a function of school level, $F_{(2, 57)} = 29.81$, $\eta_p^2 = 0.51$, $p < 0.001$. First graders were more successful (78%) than kindergartens (43%), and preschoolers (32%). Furthermore, the percentages of correct responses decreased as a function of numerosities, $F_{(2, 114)} = 98.49$, $\eta_p^2 = 0.63$, $p < 0.001$ (with 87, 38, and 28% for small, medium, and large numerosities, respectively). Moreover, there was an interaction between the two factors, $[F_{(4, 114)} = 7.89$, $\eta_p^2 = 0.22$, $p < 0.001]$ showing that the effect of school level increased with numerosities.

Finally, a correlational analysis between the transformed percentages of spontaneous use of fingers in the counting task and the transformed percentages of correct responses in this task revealed that these two variables were positively related ($r = 0.74$, $p < 0.001$). This held true when the age of children was entered as a covariate in the analysis ($r = 0.51$, $p < 0.001$). These results suggest that using fingers in the picture counting task is a good strategy that helps children keeping track of the number of pictures named. This is largely confirmed by a more descriptive observation showing that children who used their fingers in the task succeeded in 73% of the trials that were constructed with large numerosities whereas children who did not use their fingers succeeded in only 2% of the trials.

FINGER NUMERICAL PATTERN RECOGNITION
Comparison of finger numerical pattern
In order to draw our conclusions from a reliable measure and to eliminate any speed/accuracy trade-off effects, composite Z scores between accuracy and reaction times were calculated for each

participant (see **Table 3** for mean accuracy and RTs). Then, a 3 (School level: Preschoolers, Kindergartens and First graders) × 3 (Configuration: Canonical vs. Non-canonical vs. Mixed) ANOVA with the first factor as a between measure and the second factor as a repeated measure was performed on composite scores in the comparison of finger numerical pattern task.

First graders and kindergartens performed better than preschoolers $[F_{(2, 57)} = 6.51$, $\eta_p^2 = 0.19$, $p = 0.003]$. Moreover, children were more successful when comparing canonical than non-canonical configurations $[F_{(1, 57)} = 17.97$, $\eta_p^2 = 0.24$, $p < 0.001]$ or mixed configurations $[F_{(1, 57)} = 8.84$, $\eta_p^2 = 0.13$, $p = 0.004]$. There was no difference between these last two conditions $[F_{(1, 57)} = 2.69$, $p = 0.11]$. Moreover, there was no interaction between these two factors ($F < 1$). Finally, there was no correlation between the transformed percentages of spontaneous use of fingers in the counting task and the performance in the comparison of finger numerical patterns ($r = -0.18$, $p > 0.05$).

Identification of finger numerical pattern
As for the previous analysis, composite Z scores between accuracy and reaction times were calculated for each participant (see **Table 4** for mean accuracy and RTs). A 3 (School level: Preschoolers, Kindergartens, and First graders) × 2 (Configuration: Canonical vs. Non-canonical) ANOVA with the first factor as a between measure and the last factor as a repeated measure was performed on the composite Z scores in the identification of finger numerical pattern task.

First graders performed better than kindergartens $[F_{(1, 57)} = 11.56$, $\eta_p^2 = 0.17$, $p = 0.001]$, who, in turn, performed better than preschoolers $[F_{(1, 57)} = 17.77$, $\eta_p^2 = 0.24$, $p < 0.001]$. Moreover, canonical configurations led to better performance than non-canonical configurations, $F_{(1, 57)} = 18.83$, $\eta_p^2 = 0.25$, $p < 0.001$. However, planned comparisons showed that this

Table 1 | Percentages of spontaneous finger use in the counting task, as a function of school level and numerosities.

	Small	Medium	Large
First graders	41.67	83.33	85.00
Kindergartens	5.00	15.00	15.00
Preschoolers	5.00	5.00	5.00

Table 2 | Percentages of correct responses in the counting task, as a function of school level and numerosities.

	Small	Medium	Large
First graders	95.00	68.33	70.00
Kindergartens	90.00	28.33	10.00
Preschoolers	75.00	18.33	3.33

Table 3 | Percentages of correct responses and reactions times (in ms) in the comparison of finger numerical pattern task as a function of school levels and finger configurations.

	Non-canonical		Mixed		Canonical	
	Percentages	RTs	Percentages	RTs	Percentages	RTs
First Graders	91.25	2719	88.13	2574	92.50	2289
Kindergartens	85.00	2626	81.88	2802	88.13	2000
Preschoolers	83.75	3088	72.50	3274	80.63	2837

Table 4 | Percentages of correct responses and reaction times (in ms) in the identification of finger numerical pattern task as a function of school levels and finger configurations.

	Canonical		Non-canonical	
	Percentages	RTs	Percentages	RTs
First graders	98.13	2407	90.63	3175
Kindergartens	87.81	3269	82.50	4047
Preschoolers	72.19	4423	68.44	5189

was true only for first graders and kindergartens [$F_{(1, 57)} = 12.17$, $\eta_p^2 = 0.18$, $p < 0.001$ and $F_{(1, 57)} = 9.06$, $\eta_p^2 = 0.14$, $p = 0.004$, respectively] but not for preschoolers [$F_{(1, 57)} = 1.04$, $p = 0.31$].

Finally, a correlational analysis between the transformed percentages of spontaneous use of fingers in the counting task and performance in the identification task showed that these two variables were related ($r = -0.64$, $p < 0.001$). However, this correlation did no longer appear once the age of children was entered as a covariate in the analysis ($r = -0.15$, $p > 0.05$). Therefore, the relationship between the use of fingers and the identification performance of finger pattern was merely due to children's natural development.

ENUMERATION TASK (GIVE-*N* TASK)

No overt finger counting was observed in the enumeration task and the analysis was carried out on the percentages of correct responses after the arcsin transformation was applied to the data. An ANOVA with the same design as before was carried out on the transformed data (**Table 5**). The main effect of School level was significant, $F_{(2, 57)} = 24.98$, $\eta_p^2 = 0.47$, $p < 0.001$, and showed that first graders (98%) were more successful than kindergartens (73%) and preschoolers (62%), $F_{(1, 57)} = 24.76$, $\eta_p^2 = 0.30$, $p < 0.001$. Moreover, there was a main effect of Numerosities, $F_{(2, 114)} = 53.17$, $\eta_p^2 = 0.48$, $p < 0.001$, showing that children were more successful with small (98%) than medium numerosities (76%), $F_{(1, 57)} = 17.43$, $\eta_p^2 = 0.23$, $p < 0.001$ and more successful with medium than large ones (57%), $F_{(1, 57)} = 43.97$, $\eta_p^2 = 0.44$, $p < 0.001$. Furthermore, the interaction between the two variables was significant, $F_{(4, 114)} = 10.30$, $p < 0.001$, $\eta_p^2 = 0.27$ and revealed that the effect of numerosities decreased as a function of school level.

Finally, a correlational analysis between the transformed percentages of spontaneous use of fingers in the counting task and the transformed percentages of correct responses in the Give-*N* task showed that these two variables were positively related ($r = 0.59$, $p < 0.001$). Importantly, this correlation was still significant when the age of participants was entered as a covariate ($r = 0.25$, $p = 0.05$). This attests that a part of performance in the Give-*N* is related to the frequency of spontaneous finger use in counting and not only to natural development. However, it is crucial to note that, on a more descriptive level, 5 children out of the 21 who scored the highest on the Give-*N* task were children who did not use their fingers in the picture counting task. This attests that using fingers to count is not a necessary stage for developing good verbal numerical abilities.

Table 5 | Percentages of correct responses in the Give-*N* task as a function of School level and Numerosities of collection.

	Small	Medium	Large
First Graders	100.00	98.33	95.00
Kindergartens	98.25	73.68	45.61
Preschoolers	96.67	56.67	31.67

GENERAL DISCUSSION

The aim of this study was to clarify the nature of the relationship between finger use and numerical processing. We were interested in determining whether finger use constitutes a necessary step for the development of numerical abilities. If finger use constitutes a necessary transient developmental stage between non-symbolic and symbolic numerical abilities, children who do not use their fingers for counting should not be able to succeed in numerical tasks. Less drastically, if finger use is rather a useful tool for the development of later numerical abilities, children who use their fingers more frequently should be more successful in numerical tasks than children who use their fingers more rarely. In order to determine which of these alternatives has to be retained, we assessed 4–7-year-old-children's performance on their spontaneous use of fingers in a counting task, on a finger numerical pattern recognition task and on an enumeration task (i.e., Give-*N* task). Then, and crucially for our purpose, we examined the correlations between the frequencies of finger use in the counting task with the performance on the other numerical tasks under study.

We showed that the percentages of spontaneous finger use in a counting task increased with school level. In fact, out of 20 children in each age group, only 1 preschooler and 3 kindergartens used their fingers, whereas 18 first graders used the finger strategy to solve the task. As expected, finger use was obviously a good strategy in this task because its frequency positively correlated with the percentages of correct responses in the task, even when the age was considered as a covariate. Moreover, it was virtually impossible for children who did not use their fingers to succeed in the task when large numerosities were concerned. This strong relationship between finger use and success in the task attests that covert use of fingers or other means to represent numbers were not implemented by children. Indeed, if unnoticed external aids have been used, such trials would also have been associated to correct responses and no correlation would have been observed. This indicates that the task we conceived is a good test in order to determine whether or not children would spontaneously implement a strategy based on finger counting. Furthermore, as already mentioned in the Introduction, the picture counting task is a powerful tool to assess finger use in children because it can reveal whether finger counting belongs to the child's strategy repertoire even after she or he has ceased to use fingers for calculations. Indeed, when children do not use their fingers to solve arithmetic problems, it is impossible to determine whether they do not need them any longer or whether they never have resorted to them. On the contrary, in the picture counting task, fingers are still required to succeed in the task. Then, children who do not use their fingers are necessarily children for whom this strategy is not available. We can therefore confidently conclude that children who do not use their fingers in the picture counting task do not correspond to children who relied on their fingers in previous stages of development.

Then, we showed that finger pattern comparison and finger pattern identification do not correlate with the frequency of finger use when the age of children is neutralized. Thus, it is not because children use their fingers more frequently in a counting task that it will help them to identify canonical and non-canonical finger

patterns more accurately or more quickly than children who use their finger more rarely. Moreover, whatever their school level, children seem to be sensitive to the canonicity of configurations in the comparison task, despite the fact that preschoolers and kindergartens very rarely use their fingers to count. This attests that a mere observation or exposure to canonical configurations is sufficient to recognize them. Still, it is interesting to note that, as soon as the task requires more than a simple pattern processing, the sensibility to canonical patterns is no longer observed in younger children. It turns out that the comparison task can be performed on a pure perceptual basis because two patterns have simply to be judged as similar or different. Our results show that, whatever their age, children benefit from the familiarity of canonical configurations. Nevertheless, the identification task not only requires children to process the patterns but also to recognize and label them. Our results show that the matching between canonical patterns and verbal tags has not been strongly established in younger children.

Finally and crucially for our purpose, we showed that there is a correlation between the frequency of finger use in a counting task and the percentages of correct responses in an enumeration task (i.e., Give-*N* task). Because children who use their fingers more frequently are mainly the oldest children in our study, we had to ensure that this correlation was not only due to natural development. For this purpose, we entered the age of children as a covariate in our analysis and showed that the correlation was still significant. Therefore, we can conclude quite confidently that finger use is a useful tool for the development of symbolic, at least verbal, numerical abilities. However, we also showed that despite this correlation, some children who do not use their fingers in the counting task are able to perform optimally in the enumeration task. Then, and

in accordance to Crollen et al. (2011a,b), our study suggests that fingers are not a necessary tool for the development of counting abilities. In others words, using fingers could constitute a beneficial step for an efficient transition from non-symbolic to symbolic numerical skills. Yet, succeeding in manipulating symbolic numerical representations is possible without this transitional stage. Within the embodied cognition framework, our results suggest that the integration of motor input during early finger counting and finger calculation can help children in their later numerical acquisitions. However, this integration does not seem necessary to develop accurate symbolic numerical representations.

Overall and interestingly, our pattern of results show that the use of finger in a counting task is related to the ability of children in a dynamic enumeration task but not to static tasks involving recognition and comparison of finger patterns. Therefore, it could be that the link between finger and numbers remain circumscribed to counting tasks and do not extent to static finger montring situations.

Beside theoretical considerations, our results could have implications concerning educational issues, assessment of children with numerical difficulties and remediation of numerical skill impairments. Indeed, it could be fruitful to more explicitly encourage children in using their fingers and establishing the link between fingers and numerosities. This could help them in constructing stable numerical representations in strengthening the link between concrete and analog representations and verbal symbolic codes (Fayol and Seron, 2005, but see Brissiaud, 2013 for a different point of view). Moreover, our study confirms the relevance of evaluating finger use in early neuropsychological assessments of numerical skills.

REFERENCES

Andres, M., Di Luca, S., and Pesenti, M. (2008). Finger-counting: the missing tool? *Behav. Brain Sci.* 31, 642–643. doi: 10.1017/S0140525X08005578

Andres, M., Seron, X., and Olivier, E. (2007). Contribution of hand motor circuits to counting. *J. Cogn. Neurosci.* 19, 563–576. doi: 10.1162/jocn.2007.19.4.563

Arp, S., and Fagard, J. (2001). Habiletés visuo-manuelles et utilisation du doigt dans le comptage chez l'enfant IMC. *Arobase* 5, 35–42.

Baddeley, A. (1986). *Working Memory*. New York, NY: Oxford University Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–660.

Brissiaud, R. (2013). *Apprendre à Calculer à l'École: Les Pièges à Eviter en Contexte Francophone*. Paris: Retz.

Butterworth, B. (1999). *The Mathematical Brain*. London: Macmillan.

Cannard, C., Bonthoux, F., Blaye, A., Scheuner, N., Schreiber, A.-C., and Trinquart, J. (2006). BD2I: normes sur l'identification de 274 images d'objets et leur mise en relation chez l'enfant Français de 3 à 8 ans. *L'Année Psychol.* 106, 375–396. doi: 10.4074/S0003503306003034

Costa, A. J., Lopes Silva, J. B., Chagas, P. P., Krinzinger, H., Lonneman, J., Willmes, K., et al. (2011). A hand full of numbers: a role for offloading in arithmetics learning? *Front. Psychol.* 2:368. doi: 10.3389/fpsyg.2011.00368

Crollen, V., Mahe, R., Collignon, O., and Seron, X. (2011a). The role of vision in the development of finger number interactions: finger-counting and finger-montring in blind children. *J. Exp. Child Psychol.* 109, 525–539.

Crollen, V., Seron, X., and Noël, M.-P. (2011b). Is finger-counting necessary for the development of arithmetic abilities? *Front. Psychol.* 2:242. doi: 10.3389/fpsyg.2011.00242

Dantzig, T. (1962). *Number: the Language of Science, 4th Edn.* London: George Allen and Unwin.

Di Luca, S., Grana, A., Semenza, C., Seron, X., and Pesenti, M. (2006). Finger-digit compatibility in Arabic numeral processing. *Q. J. Exp. Psychol.* 59, 1648–1660. doi: 10.1080/17470210500256839

Domahs, F., Krinzinger, H., and Willmes, K. (2008). Mind the gap between both hands: evidence for internal finger-based number representations in children's mental calculation. *Cortex* 44, 359–367. doi: 10.1016/j.cortex. 2007.08.001

Domahs, F., Moeller, K., Huber, S., Willmes, K., and Nuerk, H.-C. (2010). Embodied numerosity: implicit hand-based representations influence symbolic number processing across cultures. *Cognition* 116, 251–266. doi: 10.1016/j.cognition. 2010.05.007

Fayol, M., Barrouillet, P., and Marinthe, C. (1998). Predicting arithmetical achievement from neuropsychological performance: a longitudinal study. *Cognition* 68, 63–70. doi: 10.1016/S0010-0277(98) 00046-8

Fayol, M., and Seron, X. (2005). "About numerical representations: insight from neuropsychological, experimental, and developmental studies," in *Handbook of Mathematical Cognition,* ed J. I. D. Campbell (New York, NY and Hove: Psychology Press), 3–22.

Fischer, J. P. (2010). Numerical performance increased by finger training: a fallacy due to regression toward the mean. *Cortex* 46, 272–273. doi: 10.1016/j.cortex.2008. 06.010

Fischer, M. H., and Brugger, P. (2011). When digits help digits: spatial-numerical associations point to finger counting as prime example of embodied cognition. *Front. Psychol.* 2:260. doi: 10.3389/fpsyg.2011.00260

Gerstmann, J. (1940). Syndrome of finger agnosia, disorientation

for right and left, agraphia and acalculia. *Arch. Neurol. Psychiatry* 44, 398–408. doi: 10.1001/arch-neurpsyc.1940.02280080158009

Gracia-Bafalluy, M., and Noël, M. P. (2008). Does finger training increase young children's numerical performance? *Cortex* 44, 368–375.

Kinsbourne, M., Elizabeth, B. M., and Warrington, K. (1963). The developmental Gerstmann syndrome. *Arch. Neurol.* 8, 490–501. doi: 10.1001/archneur.1963.00460050040004

Lecointre, A.-S., Lépine, R., and Camos, V. (2005). "Développement et troubles des processus de quantification," in *La Dyscalculie: Trouble du Développement Numérique de l'Enfant,* ed M. P. Noël (Marseille: Solal), 41–75.

Marinthe, C., Fayol, M., and Barrouillet, P. (2001). "Gnosies digitales et développement des performances arithmétiques," in

Troubles du Calcul et Dyscalculies chez l'Enfant, eds A. Van Hout and C. Meljac (Paris: Masson), 239–254.

Moeller, K., Fischer, U., Link, T., Wasner, M., Huber, S., Cress, U., et al. (2012). Learning and development of embodied numerosity. *Cogn. Process.* 13, S271–S274. doi: 10.1007/s10339-012-0457-9

Nicoladis, E., Pika, S., and Marentette, P. (2010). Are number gestures easier than number words for preschoolers? *Cogn. Dev.* 25, 247–261.

Noël, M. P. (2005). Finger gnosia: a predictor of numerical abilities in children? *Child Neuropsychol.* 11, 413–430.

Plaisier, M. A., and Smeets, J. B. J. (2011). Number magnitude to finger mapping is disembodied and topological. *Exp. Brain Res.* 209, 395–400. doi: 10.1007/s00221-011-2562-x

Reeve, R., and Humberstone, J. (2011). Five- to 7-year-olds' finger gnosia and calculation abilities. *Front. Psychol.* 2:359. doi: 10.3389/fpsyg.2011.00359

Rusconi, E., Walsh, V., and Butterworth, B. (2005). Dexterity with numbers: rTMS over left angular gyrus disrupts finger gnosis and number processing. *Neuropsychologia* 43, 1609–1624. doi: 10.1016/j.neuropsychologia.2005.01.009

Sandrini, M., and Rusconi, E. (2009). A brain for numbers. *Cortex* 45, 796–809. doi: 10.1016/j.cortex.2008.09.002

Thevenot, C., and Fluss, J. (2012). Numerical abilities in children with congenital hemiplegia. *Int. J. Psychol.* 47, 572–572.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

Table A1 | List of the 24 trials presented in the comparison of finger numerical pattern (adapted from Noël, 2005 and Gracia-Bafalluy and Noël, 2008).
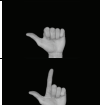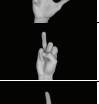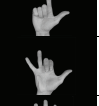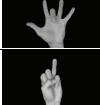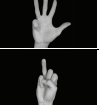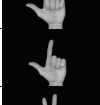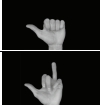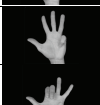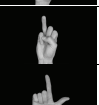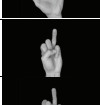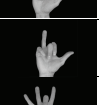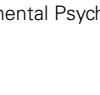
	Conditions	Left hand		Right hand
Trial 1	Same numerosity Canonical patterns			
Trial 2	Same numerosity Canonical patterns			
Trial 3	Same numerosity Canonical patterns			
Trial 4	Same numerosity Canonical patterns			
Trial 5	Same numerosity Canonical and non-canonical patterns			
Trial 6	Same numerosity Canonical and non-canonical patterns			
Trial 7	Same numerosity Canonical and non-canonical patterns			
Trial 8	Same numerosity Canonical and non-canonical patterns			
Trial 9	Same numerosity Non-canonical patterns			
Trial 10	Same numerosity Non-canonical patterns			
Trial 11	Same numerosity Non-canonical patterns			
Trial 12	Same numerosity Non-canonical patterns			
Trial 13	Different numerosities Canonical patterns			
Trial 14	Different numerosities Canonical patterns			
Trial 15	Different numerosities Canonical patterns			
Trial 16	Different numerosities Canonical patterns			
Trial 17	Different numerosities Canonical and non-canonical patterns			
Trial 18	Different numerosities Canonical and non-canonical patterns			
Trial 19	Different numerosities Canonical and non-canonical patterns			
Trial 20	Different numerosities Canonical and non-canonical patterns			
Trial 21	Different numerosities Non-canonical patterns			
Trial 22	Different numerosities Non-canonical patterns			
Trial 23	Different numerosities Non-canonical patterns			
Trial 24	Different numerosities Non-canonical patterns			

Table A2 | Additional pictures presented in the identification of finger numerical pattern task (adapted from Noël, 2005 and Gracia-Bafalluy and Noël, 2008).

Numerosity	Canonical patterns	Non canonical patterns First version	Non canonical patterns Second version
6			
7			
8			
9			

Estimation abilities of large numerosities in Kindergartners

Sandrine Mejias * and Christine Schiltz

Educational Measurement and Applied Cognitive Science, Université du Luxembourg, Walferdange, Luxembourg

The approximate number system (ANS) is thought to be a building block for the elaboration of formal mathematics. However, little is known about how this core system develops and if it can be influenced by external factors at a young age (before the child enters formal numeracy education). The purpose of this study was to examine numerical magnitude representations of 5–6 year old children at 2 different moments of Kindergarten considering children's early number competence as well as schools' socio-economic index (SEI). This study investigated estimation abilities of large numerosities using symbolic and non-symbolic output formats (8–64). In addition, we assessed symbolic and non-symbolic early number competence (1–12) at the end of the 2nd ($N = 42$) and the 3rd ($N = 32$) Kindergarten grade. By letting children freely produce estimates we observed surprising estimation abilities at a very young age (from 5 year on) extending far beyond children's symbolic explicit knowledge. Moreover, the time of testing has an impact on the ANS accuracy since 3rd Kindergarteners were more precise in both estimation tasks. Additionally, children who presented better exact symbolic knowledge were also those with the most refined ANS. However, this was true only for 3rd Kindergarteners who were a few months from receiving math instructions. In a similar vein, higher SEI positively impacted only the oldest children's estimation abilities whereas it played a role for exact early number competences already in 2nd and 3rd graders. Our results support the view that approximate numerical representations are linked to exact number competence in young children before the start of formal math education and might thus serve as building blocks for mathematical knowledge. Since this core number system was also sensitive to external components such as the SEI this implies that it can most probably be targeted and refined through specific educational strategies from preschool on.

Keywords: approximate number system, early number competence, numeracy, estimation, non-symbolic numbers, symbolic numbers, socio-economic factors, mathematical development

INTRODUCTION

Math abilities are of fundamental importance in modern society and possessing good mathematical knowledge critically determines the likelihood of employment (e.g., Rivera-Batiz, 1992). Yet we are unfortunately not all equal in learning math: Some of us excel in the mathematical domain and dedicate their careers to it while others struggle with in school and later avoid it at any cost. But even before formal math education has started young children do not enter school with the same chances. Especially pupils from low socio-economic families seem to be at risk for mathematical failure and a difference in early number skills was already noticed in preschoolers (e.g., Jordan et al., 2006) which was then evolving toward a global mathematics underachievement in middle to high school students (Dossey et al., 1988).

In all these cases, however, math ability is thought to develop based on the Approximate Number System (ANS), an ontogenetically and phylogenetically primitive system dedicated to numerosity processing (Cantlon, 2012). The ANS is known to develop throughout the lifespan. Yet, how factors such as education and socio-economic environment influence this development are fundamental questions that still need to be fully elucidated. To what extend ANS serves as a building block for

arithmetical knowledge and supports procedures for numerical computation is a related issue. Indeed, understanding the basis of typical development will help us developing good educational strategies, identifying the deficits observed in mathematical learning difficulties and dyscalculia and elaborating evidence-based guidelines for remediation.

Up to now, three categories of behavioral tasks have been used to assess approximate number representations in animals and humans: estimation, comparison, and approximate calculation. Performance in those tasks is supposed to index specifically the memory [1] representations of the analog quantity system and is thought to reflect the quality of an individual's ANS. Several regularities across the different types of comparison, approximate calculation, or estimation tasks could be singled out.

What are the signatures of the ANS and its development? In *comparison* (i.e., choose the largest numerosity) and *approximate*

---

[1] That is representations in the sense of first and second order isomorphisms (the structure of the representation contains information about the structure of the object that is represented and the relations that hold between external objects are supposed to exist in a similar fashion in the corresponding form of mental representations, e.g., Lass et al., 1993).

calculation tasks (i.e., add or subtract two large numerosities and compare the resulting sum to a third numerosity), participants' performance depends on the numerical ratio between the non-symbolic stimuli. This corresponds to a limit of the system, which can be measured through the Weber fraction (W), an important signature of the ANS. Up to now research in typical development has consistently revealed that the critical discrimination ratio narrows with age, i.e., the ability to discriminate between two numerosities improved with age (see Halberda et al., 2012, for the reverse trend in elderly). This ability, already present a few hours after birth (Izard et al., 2009), allows infants to discriminate the numerosity of small sets of objects (e.g., Starkey and Cooper, 1980), or even larger ones when the ratio between them is large enough (e.g., Xu and Spelke, 2000). A developmental increase in precision was reported by Piazza et al. (2010) using a classical comparison task of two dot sets in 5 and 10 year old children. Mundy and Gilmore (2009) also showed this increase in another comparison paradigm: the children had to map a symbolic target (i.e., Arabic symbols presented with pre-recorded number words) with one of the two alternative non-symbolic choices (dot sets of 20–50 dots) or they had to do the reverse mapping (map a non-symbolic target numerosity with one of the two alternative symbolic choices). The authors observed a performance increase between 6 and 8 years of age with generally better results for the mapping from dots to Arabic numbers than for the reverse.

Performance in *estimation tasks* (i.e., freely produce a symbolic or non-symbolic equivalent of the numerosity) confirm that, non-human species (e.g., Platt and Johnson, 1971) and humans are able to process numerical quantities approximately, no matter what the modality/format of the input and output are (non-symbolic to symbolic visual or verbal, Whalen et al., 1999; Castronovo and Seron, 2007 or the reverse symbolic to non-symbolic mapping process, Whalen et al., 1999; Cordes et al., 2001). In these tasks, over- and underestimation errors are observed, which seem to depend on the direction of the mapping, in the sense that overestimations are associated with symbolic to non-symbolic mappings, whereas non-symbolic to symbolic mapping is related to underestimations (see for example Castronovo and Seron, 2007; Crollen and Seron, 2012). Moreover, the estimations of the target magnitudes are generally inaccurate such that mean estimates and response variability both increase with target magnitude, indicating that the underlying representation is less precise for larger numerosities. More specifically, this representation is characterized by a scalar variability which gives rise to a constant coefficient of variation (COV = standard deviation of mean response/mean response) across target magnitudes (Whalen et al., 1999; Cordes et al., 2001; Mejias et al., 2012a,b; Castronovo and Göbel, 2012). Over early development, the variability of the representations decreases with age while their precision increases. Indeed, studies of typical development have consistently reported increasing precision of the ANS with age. In a study by Huntley-Fenner's (2001), 5–7 year olds had to estimate the numerosity of a set of black squares (5–11 items) on a number line consisting of a series of Arabic numbers ordered from 0 to 20. Mean accuracy significantly increased throughout the age range of 5–7 years and COV scores were negatively correlated with age in days (COVs ranged from 0.37 to 0.11), showing

that estimates were less variable with increasing age. In an estimation study using larger numerosities, Mejias et al. (2012b) reported that mean 9 year olds' COV was 0.29, which is higher than adult's mean COV (0.16 in Mejias et al., 2012a). The precision and the variability of this approximate analog representation consequently seem to be related to development (see also Chillier, 2002; Booth and Siegler, 2006).

How do ANS refinements relate to math achievement? Several studies directly investigated the link between children's ANS and their abilities in learning numbers symbols and arithmetic. Since study outcomes were quite divergent, it is not yet clear how the interactions between ANS and symbolic number knowledge arise and develop. The conflicting outcomes could have different origins because different age-ranges of populations have been tested with different types of tasks (i.e., comparison, approximate calculations, and estimation as detailed above) probing ANS.

In Piazza and collaborators' study (2010), 5–10 year old children's non-symbolic number acuity did not correlate with (symbolic) arithmetical scores (for similar results in 4–7 year olds see Soltesz et al., 2010). In Mundy and Gilmore's study with 6–8 year olds, scores on the mapping tasks did not correlate with arithmetical scores. Congruent with those results, the school mathematics performance of 6–8 year old children was found to be unrelated to the magnitude of their numerical distance effect exhibited in a comparison task involving non-symbolic numerical displays (Holloway and Ansari, 2009; Sasanguie et al., 2013). Such a relation was on the contrary obtained on a similar comparison task with symbolic numbers in 6 year olds, since children with a smaller symbolic distance effect showed higher mathematics performance (De Smedt et al., 2009; Holloway and Ansari, 2009).

Yet, others did, however, report that western adolescents' ANS precision was clearly related to performance in exact calculation and number processing. They observed a correlation between the accuracy in a non-symbolic numerical comparison task performed at age 14 and school mathematics performance from Kindergarten to sixth grade (Halberda et al., 2008). Gilmore et al. (2010) also provide evidence that the ANS precision might be related to symbolic knowledge. Based on studies of Barth et al. (2005, 2006) they evaluated 5–6 year old children during their first year of school and found that children's performance on large-number non-symbolic approximate addition related significantly to their mastery of school's mathematics curriculum at the end of that first year of formal instruction (e.g., counting objects, recognizing Arabic digits, symbolic and non-symbolic comparisons—all numerical tasks using numbers smaller than 10). It appeared that non-symbolic arithmetic performance was related to children's mathematics achievement 3 months later, independently of achievement levels in reading or intelligence and socio-economic background. Congruently, Mussolin et al. (2012) examined the performance of 3 different groups of children from Kindergarten to grade 1. When considering them together, a positive relation between their accuracy in discriminating sets of non-symbolic elements and their ability to process numerical symbols was observed even when taking intelligence and short-term memory into account (see also Libertus et al., 2011). Finally, data collected from Amazonian Munduruku indigene children

and adults show that Munduruku with a certain level of symbolic number knowledge have a more refined ANS than their peers with little or no formal instruction (Pica et al., 2004). Similarly, it was recently reported that math education sharpens the approximate numerical representations in western adults (Nys et al., 2013).

In summary, the ANS seems to serve representing the approximate cardinal values of large sets of stimuli and it can be assessed using estimation, comparison and approximate calculation tasks. Parallel signatures of the ANS were found in studies of human adults, children, infants (and even non-human animals). They provide evidence for the existence of a magnitude-based estimation system for representing symbolic and non-symbolic numerical magnitude that also supports procedures for numerical computation, even outside formal education. According to some authors (e.g., Barth et al., 2005), this is congruent with the fact that the ANS serves as building block for symbolic arithmetic learning. However, currently this assumption is also challenged by several studies, which consistently failed to observe a correlation between non-symbolic numerical magnitude comparison and mathematical performance at the beginning of formal school education (e.g., Holloway and Ansari, 2009; Mundy and Gilmore, 2009). To resolve this contradiction and better understand the observed changes in ANS acuity, further studies investigating young children's relation between ANS and exact number knowledge are urgently required.

In the present study, we evaluated preschool children's ability to estimate large non-symbolic numerosities. Children had to produce estimates of large numerosities ranging from 8 to 64 elements, i.e., clearly exceeding number values included in their school curricula [2]. Estimation abilities for these large quantities were assessed via both non-symbolic (i.e., 64 differently sized elements) and symbolic (i.e., "64") output formats. To the best of our knowledge there are currently no studies investigating large quantity estimation in preschoolers. The rational was to assess and compare both symbolic as well as non-symbolic estimation abilities of large numerosities (and the underlying ANS representations) in preschool children, i.e., before these numerosities are systematically learned and their exact meaning is mastered. To highlight the influence of preschool math education on estimation abilities we compared children from the 2nd and the 3rd Kindergarten grade while they were performing symbolic and non-symbolic estimation. In addition children's early number competence levels associated with the two Kindergarten grades were evaluated using exact number processing tasks involving numerosities up to 12 items. To analyze the effect of environmental influences such as the socio-economic status on approximate (and exact) number abilities in preschool children, we compared the performances of children coming from two schools characterized by different levels of socio-economic index (SEI). Finally, we used an individual differences approach to investigate whether preschoolers' early (symbolic and non-symbolic) number competence might relate to their accuracy in these approximate numerical tasks.
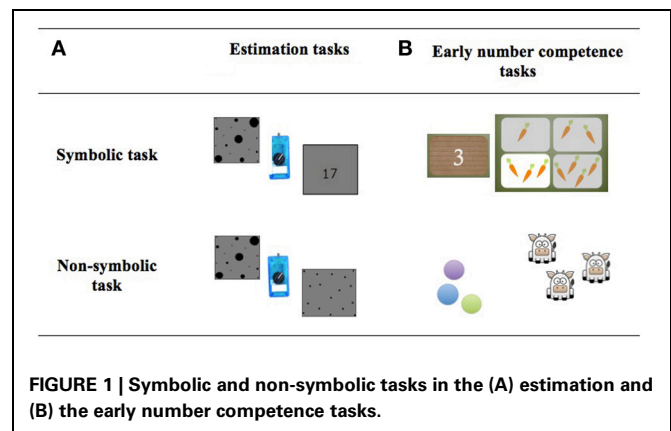
FIGURE 1 | Symbolic and non-symbolic tasks in the (A) estimation and (B) the early number competence tasks.

Because our design included both symbolic and non-symbolic tasks we could systematically investigate the relationship between exact and approximate numerical abilities for these two task formats (see **Figure 1**). While others have evaluated either symbolic (i.e., Gilmore et al., 2007; Mundy and Gilmore, 2009) or non-symbolic (i.e., Gilmore et al., 2010) approximate number abilities, there are currently no studies evaluating the two types of approximate processing within the same population of preschool children (see (Mejias et al., 2012b) for this type of evaluation in 9–10 year old 4th graders). This seems, however, particularly important given the above-mentioned controversies concerning the role of non-symbolic vs. symbolic number abilities as precursors of math performance (e.g., Halberda et al., 2008 vs. De Smedt et al., 2009; Holloway and Ansari, 2009; Gilmore et al., 2010). Indeed it is complicated to compare the correlations between approximate numerical abilities and math competence observed in different studies since they might be confounded by subtle differences related to group (e.g., age, environmental context) or study design (e.g., math test battery).

In addition we believe that the free estimation tasks used in the present study are more sensitive to individual differences in numerical processing than the predominantly used comparison tasks. First, the comparison paradigms might lack sensitivity because they assess performance for a limited number of predefined numerical ratios (e.g., Rousselle and Noël, 2007; De Smedt and Gilmore, 2011). Under those conditions, it is always possible to miss a significant difference between participant groups if the ratios selected are not sensitive enough. Second, it was recently argued that the mechanisms used to extract information from dot-arrays in comparison tasks is driven by visual features rather than numerical dimensions (Szucs et al., 2013). In contrast, numerical estimation tasks cannot be solved by only relying on perceptual processes since they typically require producing estimation outputs in a different format than the input (e.g., heterogeneous to homogeneous dots, dots to Arabic digits, dots to number words). Consequently we preferred to use a free production paradigm in which children have to produce a certain magnitude. This allowed us to measure directly the precision of the children's estimates. Accordingly, we hoped to unveil the so far controversial relationship between the ANS and children's early number competence.

---

[2]In Belgium, by the end of the Kindergarten, children are familiarized and manipulated numbers until 6.

We hypothesized that preschoolers would show the ANS signature when producing symbolic and non-symbolic estimation outputs for numbers largely exceeding their curricular pre-mathematical knowledge. Moreover, 3rd grade preschool children were expected to be more accurate than their peers from the 2nd grade. Especially the non-symbolic tasks assessing approximate and exact numerical abilities were expected to co-vary from an early age, whereas a certain level of number symbol mastery might be required before exact number symbols are linked to ANS. Concerning the impact of school's SEI on estimation abilities, predictions were less clear-cut given the mixed evidence in the literature (Ramani and Siegler, 2008; Gilmore et al., 2010). Yet in any case, it is critical to identify if and when external factors such as SEI impact estimation abilities (and early number competence) in order to optimally design and plan educational intervention.

METHODS

PARTICIPANTS

Participants were 74 children coming from two different Belgium public Kindergarten schools. Parental consent was obtained for each of the children. One school was ranked as a school with a low SEI[3] whereas the other was a middle SEI school. In each school, one group of children was tested at the end of the second grade (4–5 year olds, mid-June 2012) and another group at the end of the third grade (5–6 year olds, mid-June 2012).

---

[3]The SEI of schools was established in Belgium in 1998 to allocate resources within the framework of the positive discrimination. This index is updated every five years and it is constructed from the variables "per capita income, educational attainment, unemployment, occupational and comfort level of housing." To each student corresponds an index defined by its area of residence. It is the smallest administrative unit for which socioeconomic data are available. The SEI is then defined based on the average of the indices of its student population; it does not correspond directly to the area of implantation, or a measure of school performance. It allows one to rank schools on a scale of 1–20, from the lowest SEI to the highest. The choice of variables, indices and formula has been approved by the Government of the French Community (de Villers and Desagher, 2011).

Children's descriptive information, according to the Kindergarten class they belonged to and their school's SEI, are presented in **Table 1A**.

Children who took part in the study had no history of developmental disorders and were considered as typically developing children by the Belgian psycho-medico-social services.

MATERIALS AND PROCEDURE

Estimation tasks

Two computerized estimation tasks developed by Mejias et al. (2012b) were used to evaluate the children's ability to estimate large quantities. They took place on a PC-compatible portable (screen size: 30.5 × 23 cm) running E-Prime software (Schneider et al., 2002). Children sat about 55 cm away from the computer screen and had to estimate the numerosities of black dots displayed for 1 s on a gray screen in two tasks: (1) In the *symbolic estimation task*, children were presented with a set of heterogeneously-sized dots. They were asked to estimate the cardinality of each set by producing the corresponding Arabic number (AN, Arial font with a visual angle of 2.2°) using a potentiometer. In each set the size of the dots was manipulated so the total covered area was identical. However, to avoid larger collections also being those with smaller elements, dots of the smallest and largest size (respectively, with visual angle of 0.44 and 0.88°) were included in all sets. (2) In the *non-symbolic estimation task*, the children were presented with the same kind of sets of dots of mixed sizes but had to produce a collection of approximately the same number of equally sized dots (see **Figure 1A**).

Four numerosities were presented (8, 16, 34, and 64) six times to the children, providing a total of 24 stimuli for each task. Four practice trials by task were proposed to the children to familiarize themselves with the experimental setting on other numerosities (15, 25, 50, 75). In this training session, the participants received feedback (on the computer screen) corresponding to the correct answer, in order to allow a calibration of their estimation (Izard and Dehaene, 2008). Data from these trials are not reported in the analyses.

Table 1 | (A) Descriptive information; (B) means (*SD*) of children's precision of numerical estimation calculated as an Absolute Error Score (AES; computed as: |the child's estimate answers—target magnitude|) by task for the different groups of children according the time of testing and the school SEI; (C) means (*SD*) of children's number of corrects trials by task for the different groups of children according the time of testing and the school SEI.

Time of testing—*School SEI*	2nd grade—*low*	2nd grade—*middle*	3rd grade—*low*	3rd grade—*middle*
(A) DESCRIPTIVE INFORMATION				
N (boys)	20 (10)	22 (11)	14 (5)	18 (9)
Age in months (*SD*)	59.15 (3.92)	57.82 (3.66)	71.79 (5.18)	70.78 (3.21)
(B) ESTIMATION TASKS				
Symbolic estimation task (*SD*)	71.08 (47.50)	53.39 (35.44)	45.64 (40.60)	32.23 (21.30)
Non-symbolic estimation task (*SD*)	52.08 (41.75)	37.62 (27.11)	35.08 (26.92)	15.40 (5.46)
(C) EARLY NUMBER COMPETENCE TASKS				
Symbolic association task (*SD*)	4.45 (2.35)	5.96 (3.89)	9.50 (2.88)	10.72 (1.27)
Non-symbolic trade task (*SD*)	5.20 (3.17)	6.23 (2.40)	9.36 (2.53)	10.67 (1.78)
Total early number competence score (*SD*)	9.65 (4.72)	12.18 (5.74)	18.86 (5.07)	21.39 (2.66)

Early number competence tasks

To assess counting development, two tasks were administered to each child individually.

1. The symbolic *association task* tested the children's knowledge of Arabic digit symbols from 1 to 12 with their associated quantities. This task was administered in three steps starting with one card (A4 format) representing four sets of vegetables (from 1 to 4), followed by another card representing 4 other sets of vegetables (from 5 to 8) and the last card with 4 larger numerosities (from 9 to 12). Children were asked to associate the set of four small cards showing Arabic digits (from 1 to 4, 5 to 8, and 9 to 12 for the three steps, respectively) with the "right quantity of vegetables." Trials were administered in a non-consecutive order within each step until the child made errors on three consecutive numerosities. The "association score" correspond to the highest number of correct association succeeded by the child.

2. The non-symbolic *trade task* assessed the non-verbal understanding of the one-to-one correspondence principles and was similar to the one used by Rousselle and Noël (2008). Children were given tokens while the experimenter presented them toy animals. In this context they were then invited to buy animals with their tokens: "You have coins and I have animals to sell. You can buy ONE animal by giving me ONE coin, no matter the size of the coin." Two practice trials ensued in which the young participant was encouraged to exchange one large and then one small coin for one animal. Those practice trials were done to make sure that the children understood that coins' size did not matter. Then, three blocks were administered: starting with block A using numerosities from 1 to 4, then block B with numerosities from 5 to 8, and finally block C with numerosities ranging from 9 to 12. Numerosities inside a block were presented in a non-consecutive order, e.g., 2, 1, 4, and 3. Children were allowed to perform the task by giving one token for one animal or by counting. The task was interrupted if the child made errors on three consecutive numerosities (e.g., if a child failed on 5, 6, and 7). The "trade score" corresponded to the highest number of animals correctly traded (in this e.g., the trade score will be 4; see **Figure 1B**).

RESULTS

ESTIMATION TASKS

The signatures of the approximate number system

We first examined if preschool children showed the typical signatures of the ANS. All children's mean estimates and standard deviations (SD) increased in direct proportion to the target magnitudes while the coefficients of variation (COV; i.e., the ratio of the standard deviation to the mean estimate) remained constant across targets (**Figure 2**). In the non-symbolic and the symbolic estimation tasks the slopes of the mean estimates and their standard deviations were close to 1 (see **Tables 2A,B**), confirming the linear increase with the target size, sign of a typical numerical magnitude representation (e.g., Crollen et al., 2011; Mejias et al., 2012b). Moreover, as measured by the COV (see **Table 2C**), the variability of estimates was relative to target size in the two grade groups and in both estimation tasks: The slope of the best

linear fit to the mean COV scores did not differ from 0 ($ps > 0.1$), except for the 2nd Kindergarten-graders in the symbolic "Dots to AN" task (children showed less variability in their answer for largest magnitudes to be estimated). The COV ranged from 0.31 to 0.89, with an average value of .58 (±0.13) and from 0.20 to 0.92, with an average value of .56 (±0.17) in the present population of 5–6 year old children, respectively for the symbolic and the non-symbolic tasks. This provides direct evidence for scalar variability in preschool children's representation of numerosity in both tasks.

Children of both testing times (i.e., 2nd and 3rd grade of Kindergarten) overestimated the numerosity of the arrays (**Figure 3**). To describe this tendency we computed the response-bias [RB = (child's response − target magnitude)/target magnitude] and tested it against zero using *t*-tests. A RB of zero indicates that estimates were accurate, a negative RB that target magnitudes were underestimated and a positive RB that target magnitudes were overestimated. Contrarily to the expected underestimation predicted by the bi-directional mapping hypothesis (e.g., Castronovo and Seron, 2007), preschool children overestimated target magnitude in the symbolic "Dots to AN" task [2nd grade children RB: $M = 3.237$; $SD = 3.002$; $t_{(41)} = 6.989$, $p < 0.001$; 3rd grade children RB: $M = 1.445$; $SD = 1.955$; $t_{(31)} = 4.182$, $p < 0.001$]. They also overestimated in the non-symbolic estimation task [2nd grade children RB: $M = 2.075$; $SD = 2.024$; $t_{(41)} = 6.645$, $p < 0.001$; 3rd grade children RB: $M = 0.786$; $SD = 1.359$; $t_{(31)} = 3.270$, $p = 0.003$]. This positive RB was shown by the preschool children of the 2nd grade on every target magnitudes of both estimation tasks. Preschool children of the 3rd grade also overestimated numerosities of all target magnitudes in the symbolic "dots to AN" task. But in the non-symbolic task only the two smallest target magnitudes were overestimated (see **Figure 3**).

In summary, preschool children's stable COVs indicate that they were able to produce approximate estimates of large numerosities, which they consequently overestimated, as revealed by their positive response biases.

Whereas the COV measures how consistently children execute the estimation task with respect to the target numerosity, they do not inform about the precision of children's representations. A look at the mean estimates produced by the children indicates that their approximate non-symbolic estimations ("how many dots create a corresponding quantity") as well as their symbolic estimations ("which AN describes a corresponding quantity") were quite far from the expected target sizes. The following analyses will provide more information about children's absolute accuracy to perform symbolic and non-symbolic numerical estimation tasks.

The precision of estimates

In a second step the precision of children's numerical estimation was calculated as an absolute error score (AES) computed as follows: |participant's estimate answers—target magnitude|. The absolute value of the sum was provided as a measure of overall accuracy without paying attention to the direction of the difference between the target and the response.

FIGURE 2 | Children's mean estimates and standard deviation (SD) increased linearly with the target magnitude in the symbolic and the non-symbolic estimation tasks. The coefficients of variation (COV)—the ratios of the standard deviation to the mean—were approximately constant across the target magnitudes in both groups.
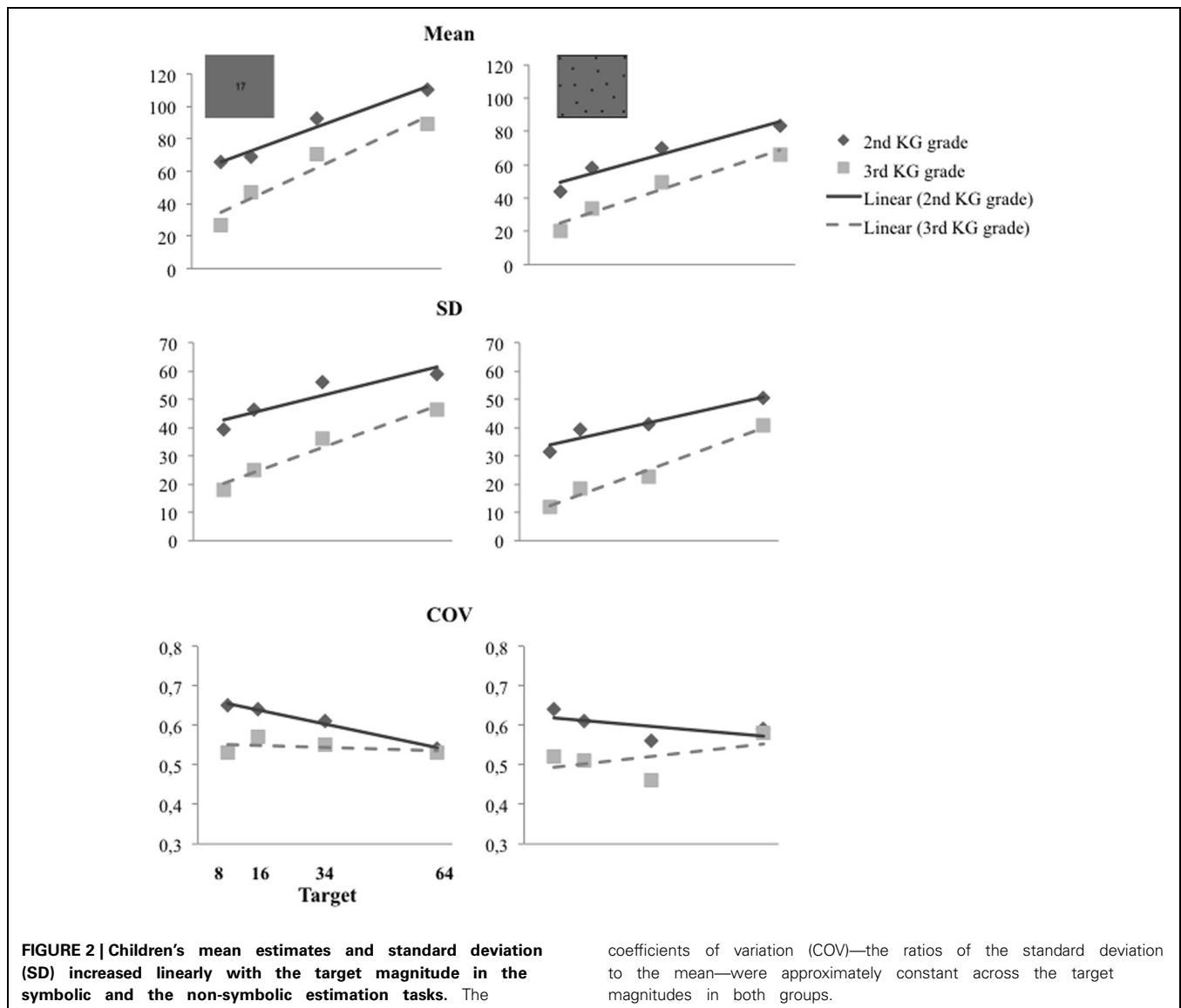
The effect of grade on the precision of estimates. To evaluate the influence of early schooling (2nd and 3rd Kindergarten grade) on the precision of numerical estimation, an ANOVA on AESs was performed with target size (8, 16, 34, 64) and tasks (non-symbolic and symbolic estimations) as within-subject factors, and the testing times (2nd and 3rd Kindergarten grade) as the between-subjects factor. According to the previous analyses regarding the scalar variability of the estimates, the target size effect was significant, $F_{(3, 216)} = 23.042$, $\eta^2 = 0.242$, $p < 0.001$, indicating that precision decreased with increasing target magnitudes. A significant effect of the task was also found, $F_{(1, 72)} = 17.209$, $\eta^2 = 0.193$, $p < 0.001$, revealing that the non-symbolic estimation task ($M = 35.643$, $SD = 31.26$) led to higher accuracy (i.e., lower AESs) compared to the symbolic estimation task ($M = 51.557$, $SD = 39.408$). Finally, the time of testing effect was significant, $F_{(1, 72)} = 9.875$, $\eta^2 = 0.121$, $p = 0.002$: 2nd grade preschool children were less accurate ($M = 53.159$, $SD = 34.176$)
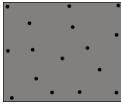
than 3rd grade children ($M = 31{,}053$, $SD = 23{,}296$). No other effect or interaction was significant.

The effect of school's SEI on the precision of estimates. To examine how important external factors such the socio-economic environment (here corresponding to the school's SEI) influence the precision of children's ANS, an ANOVA with the four target sizes, the two estimation tasks as within factors and the two levels of school' SEI as the between factor was performed for 2nd and 3rd grade children on AES (see also the descriptive information reported in **Table 1B**).

For the *2nd grade preschool children*, the target size effect was present, confirming the onset of typical approximate number representation characteristics as soon as 4–5 year olds, $F_{(3, 120)} = 7.312$, $\eta^2 = 0.155$, $p < 0.001$. The effect of task was also significant, $F_{(1, 40)} = 9.228$, $\eta^2 = 0.187$, $p = 0.004$. The non-symbolic estimation task gave rise to higher accuracy levels

Table 2 | Results of the linear regression between the predictor variable (target results) and (A) the mean of the estimates, (B) the standard deviations of the estimates, and (C) the coefficients of variation (COV) of the estimates in the two estimation tasks for the two grades tested.

2nd KG grade	3rd KG grade	2nd KG grade	3rd KG grade
(A) MEAN OF ESTIMATES			
$r^2 = 0.972$	$r^2 = 0.927$	$r^2 = 0.929$	$r^2 = 0.955$
$\beta_0 = 0.986$	$\beta_0 = 0.963$	$\beta_0 = 0.964$	$\beta_0 = 0.977$
$t_{(2)} = 8.391*$	$t_{(2)} = 5.048*$	$t_{(2)} = 5.126*$	$t_{(2)} = 6.525*$
(B) STANDARD DEVIATION			
$r^2 = 0.839$	$r^2 = 0.958$	$r^2 = 0.907$	$r^2 = 0.978$
$\beta_0 = 0.916$	$\beta_0 = 0.979$	$\beta_0 = 0.952$	$\beta_0 = 0.989$
$t_{(2)} = 3.231*$	$t_{(2)} = 6.755*$	$t_{(2)} = 4.416*$	$t_{(2)} = 9.435*$
(C) COV			
$r^2 = 0.988$	$r^2 = 0.038$	$r^2 = 0.295$	$r^2 = 0.293$
$\beta_0 = -0.994$	$\beta_0 = -0.196$	$\beta_0 = -0.543$	$\beta_0 = 0.541$
$t_{(2)} = -13.029**$	$t_{(2)} = -2.830$	$t_{(2)} = -0.914$	$t_{(2)} = 0.910$

*Correlation significantly different from 0 at p 0.05; ** at p 0.01.

than the symbolic task ($M = 44.507$, $SD = 35.177$; $M = 61.812$, $SD = 42.061$, respectively). No other effect or interaction was significant.

Regarding the *3rd grade preschool children*, the target size effect was present, $F_{(3, 90)} = 17.653$, $\eta^2 = 0.370$, $p < 0.001$, revealing the expected approximate number representation signature. The task effect was significant, $F_{(1, 30)} = 9.047$, $\eta^2 = 0.232$, $p = 0.005$, showing again that 5–6 year old children are more accurate in the non-symbolic estimation condition ($M = 24.008$; $SD = 20.463$) compared to the symbolic one ($M = 38.098$; $SD = 31.428$). Most importantly, also SEI impacted their numerical representation significantly, $F_{(1, 30)} = 4.410$, $\eta^2 = 0.128$, $p = 0.044$). Children from the low SEI school showed a less refined magnitude representation ($M = 40.360$; $SD = 30.503$) compared to children from the middle SEI school ($M = 23.814$; $SD = 12.299$).

In 3rd grade, significant double interactions between target magnitude and SEI, $F_{(3, 90)} = 5.750$, $\eta^2 = 0.161$, $p = 0.001$, and triple interactions between target magnitude, SEI and task, $F_{(3, 90)} = 3.241$, $\eta^2 = 0.097$, $p = 0.026$, were observed as well. The double interaction was due to the fact that middle SEI children were showing the expected effects of target magnitude increases on AES (the two tasks confounded, AES means were 7.694, 13.972, 27.074, 46.514 for targets 8, 16, 34, and 64, respectively) while low SEI children did not show this typical sign of approximate magnitude representation (27.185, 44.649, 44.9167, 44.691 for targets 8, 16, 34, and 64, respectively). Finally, the decomposition of the triple interaction by an ANOVA for each estimation task (4 target sizes × 2 SEIs) revealed that the 3rd grade low SEI preschool children had (i) poorer non-symbolic approximate representations [beside the Target size effect, $F_{(3, 90)} =$

11.076, $\eta^2 = 0.270$, $p < 0.001$, the analyses showed a SEI effect, $F_{(1, 30)} = 9.219$, $\eta^2 = 0.235$, $p = 0.005$]; (ii) and poorer symbolic estimation task performance over the small range of numerosities [beside the Target size effect, $F_{(3, 90)} = 9.642$, $\eta^2 = 0.243$, $p < 0.001$, a Target size × SEI interaction was significant, $F_{(3, 90)} = 6.053$, $\eta^2 = 0.168$, $p = 0.001$; see **Figure 4**].

In short, preschool children's estimates became less precise with increasing target magnitude, in line with the hypothesis that their response production relies on ANS recruitment. Moreover, 3rd grader's performances were generally more precise than that of their second grade peers. It was also in 3rd Kindergarten grade that the precision of children estimates were significantly influenced by their school's SEI (except for the production of large symbolic estimates).

EARLY NUMBER COMPETENCE TASKS

In addition to measuring preschoolers' estimation abilities, we also assessed children's exact early number competences using number processing tasks involving numerosities up to 12 items. Descriptive information concerning means and standard deviations for the symbolic association and non-symbolic trade tasks for each group of children (according the period of the testing session and the SEI) are reported in **Table 1C**. Firstly, we investigated the influences of (a) the time of testing and (b) the school's SEI on children's performances in two exact numerical tasks. The first allowed us to assess the influence of schooling whereas the second evaluated the effect of socio-economic environment on children's early number competence. Accordingly, a repeated-measures analysis of variance (ANOVA) with the different testing times (2) and school SEIs (2) as the between-subjects factor were conducted regarding the two exact numerical tasks (the non-symbolic trade and the symbolic association tasks as within-subject factors). Results showed a significant effect of time of testing, $F_{(1, 70)} = 67.974$, $\eta^2 = 0.493$, $p < 0.001$, with the preschool children of the 2nd grade ($M = 5.488$; $SD = 2.686$) reaching lower performance levels than the children of the 3rd grade ($M = 10.140$; $SD = 2.017$). The effect of SEI was also significant, $F_{(1, 70)} = 5.140$, $\eta^2 = 0.068$, $p = 0.026$, showing that lower SEI participants ($M = 6.721$; $SD = 3.322$) were less efficient compared to the children from the middle SEI school ($M = 8.163$; $SD = 3.253$). No other effect or interaction was significant. Thus, preschoolers' understanding of the one-to-one correspondence principle (non-symbolic trade task), as well as their cardinal understanding of small Arabic digit symbols (symbolic association task) were directly influenced by (a) the level of Kindergarten schooling and (b) the SEI of the school that they were attending.

CORRELATION BETWEEN NON-SYMBOLIC AND SYMBOLIC NUMBER KNOWLEDGE

In order to evaluate the relationship between preschool children's symbolic and non-symbolic exact numerical knowledge (i.e., based on their scores in the association and trade tasks, respectively) and their symbolic and non-symbolic approximate magnitude representations, correlation analyses were performed (**Table 3**).

FIGURE 3 | Response-bias (RB) in the symbolic and non-symbolic tasks: children from 2nd and 3rd grade of Kindergarten overestimated the numerosity of the arrays.

FIGURE 4 | Mean AES for low and middle SEI preschool children attending the 3rd grade, for each target magnitude in the symbolic and the non-symbolic estimation tasks. Children from the low SEI school were significantly less accurate when estimating large numerosities. In the symbolic estimation task this SEI-related difference did not pertain to the two largest quantities. Note: *Group differences significant at $p < 0.05$; ** at $p < 0.01$.

They revealed that, in 2nd and 3rd graders, the non-symbolic exact numerical task (i.e., the trade task), which assesses the non-verbal understanding of the one-to-one correspondence principles correlated with the non-symbolic estimation task. However, the entirely non-symbolic trade task did not correlate with the symbolic estimation task in either group.

Finally, the association task, which evaluated cardinal knowledge of number symbols up to 12, correlated with both symbolic and non-symbolic estimation of large quantities in 3rd graders only. This last result is congruent with the idea that children who present better exact symbolic knowledge are also those with the most refined ANS. However, this was true *only* for 3rd grade preschool children who were at a few months of receiving mathematics instructions. Indeed, no relation between the mastery of number symbols and estimation abilities could be found 1 year earlier in children attending 2nd Kindergarten grade.

DISCUSSION

To better understand approximate number processing and how it relates to exact number knowledge, the present paper explored young children's numerical abilities before they enter formal math education. To this aim we used a free estimate production paradigm and investigated for the first time preschooler's abilities to estimate large numerosities (ranging from 8 to 64 elements). Our data show that 5–6 year old children were able to produce estimates of large numerosities which reveal the typical ANS signature: Mean estimates and standard deviations both augmented constantly with increasing numerical target size. These results suggest that preschool children accessed the ANS in estimation tasks akin 9–10 year olds and adults in similar tasks (Castronovo and Göbel, 2012; Crollen and Seron, 2012; Mejias et al., 2012a,b).

By letting children freely produce estimates of several large numerosities we obtained direct insights into their ANS

Table 3 | Results of the Pearson correlations between the two exact numerical tasks (symbolic association and non-symbolic trade tasks) and the two estimation tasks (the non-symbolic and the symbolic one) performed by preschool children.

			Symbolic association task	Non-symbolic trade task
3rd grade	[dots]	r	−0.529**	−0.437*
	[17]	r	−0.382*	−0.208
2nd grade	[dots]	r	−0.252	−0.373*
	[17]	r	−0.153	−0.108

*Correlation significant at p < 0.05 (bilaterally); ** at p < 0.01 (bilaterally).

representations, in the sense that answers were not constrained by double choices (as typically the case in comparison tasks, e.g., Gilmore et al., 2007, 2010; De Smedt and Gilmore, 2011). Moreover, free estimation paradigms are more sensitive to individual differences than comparison tasks because they do not require the use of a limited number of numerosity ratios, i.e., difficulty levels (see for example Gilmore et al., 2010; De Smedt and Gilmore, 2011 vs. Halberda et al., 2008; Mussolin et al., 2012). Although preschoolers' estimates were relatively far from the targets, the stable COVs across targets (see Huntley-Fenner, 2001 for a detailed discussion of this measure) indicate that this was not due to a lack of compliance or an inability to perform the task. This performance pattern rather resulted from the fact that free estimation paradigms allow capturing more response nuances in a greater number of participants. Indeed, these paradigms also allow keeping the entire set of subjects for data analysis. This starkly contrasts with comparison studies, which tend to reject outliers (e.g., De Smedt and Gilmore, 2011; Sasanguie et al., 2013). However, especially these rejected participants could have highly informative analogue representations. In the present free estimation approach the data of all the tested preschool children were included in the analysis. Evaluating the stability of the COV across increasing target numerosities informed us on the consistency of preschoolers' responses while revealing their access to analog magnitude representations. Given the well-documented ANS acuity *increase* with age, the present COV averages (0.58 ± 0.13; 0.56 ± 0.17) from the symbolic and the non-symbolic tasks, respectively, are in line with the average COVs from the symbolic (0.36 ± 0.14) and non-symbolic (0.35 ± 0.12) tasks obtained in 9 year olds with a similar paradigm (Mejias et al., 2012a). Moreover, the difference in maximal estimation set size (20 vs. 64 here) explains why the 5 year olds in the study of Huntley-Fenner (2001) achieved a better COV of 0.37 than the children in the present study.

Contrary to the systematic underestimation bias observed in adults which perform approximate non-symbolic to symbolic mapping tasks (e.g., Castronovo and Seron, 2007), the young children of the present study overestimated the number of elements in the target sets. Although this result will need to be confirmed in future studies, it suggests a developmental trend from over- to underestimation. According to the bi-directional mapping hypothesis (Castronovo and Seron, 2007; Crollen et al., 2011; Crollen and Seron, 2012) adults are thought to underestimate large numerosities because they systematically map their logarithmically compressed approximate representation of the target set to its corresponding exact and linear representation of Arabic number symbolic (for an illustration, see Figure 2 in Crollen et al., 2011). Since preschoolers did not yet develop an exact linear representation of large number symbols (which will only be acquired in primary school), they need to rely on a more primitive and noisier symbolic representation of large numbers, resulting in systematic overestimation. The observation that preschool children performed better in the non-symbolic "dots to dots" task compared to the symbolic "dots to AN" task perfectly matches with this proposal. Combined with the overestimation, it indicates that young preschool children rely on qualitatively different estimation processes compared to older children and adults who received formal math education and perfectly master number symbols. Indeed, 9 year old typically achieving children as well as adults do not differ in non-symbolic vs. symbolic estimation tasks. Dyscalculic children, in contrast show the same profile than our pre-school children, that is a relatively better performance in the non-symbolic estimation task (Mejias et al., 2012a,b).

Although all tested preschooler populations relied on the ANS during approximate number processing, time of testing also had an impact on the ANS accuracy since 2nd grade preschool children were less precise compared to their 3rd grade peers in both symbolic and non-symbolic estimation tasks. In addition to instruction, the SEI of the school that children attended significantly influenced preschoolers' estimation performance, but only in 3rd grade of Kindergarten. This relatively late effect of the socio-economic status on the ANS precision contrasts with the fact that children's early number competence was already affected by school SEI in 2nd grade. The latter results confirm previous reports that 5 year old children from a low socio-economic environment have significantly worse early number skills compared to their socio-economically more advantaged peers at the end of Kindergarten (Jordan et al., 2006) and they extend them to 4 year olds (i.e., 2nd KG graders). The above-mentioned findings reinforce the proposal that the ANS is an innate system that is naturally predetermined to process numerosities and serves as a building block for the development of exact symbolic number competence (e.g., Barth et al., 2006). The newly learned number symbols then in turn positively influence the approximate numerical abilities. Within this theoretical framework the effects of SEI are expected to appear at different ages for exact and approximate number processing since external factors such as SEI will *first* affect the culturally acquired early number skills which will *then* through retro-influence refine the innate number sense.

In the present study, even preschool children were able to use complicated number symbols (>5 or even 10) to estimate numerosities in a non-random and meaningful way characterized

by ANS signatures. However, in the youngest group these approximate symbolic abilities were completely independent of exact symbolic number. The latter link did, eventually, emerge in 3rd Kindergarten grade. Just as Sasanguie et al. (2013) observed with young primary school children, 3rd grade preschoolers who showed the best early number competence were also those who could process non-symbolic numerical magnitudes most precisely in the present study. The process of mapping exact symbolic knowledge onto the innate non-symbolic system thus seems to start even sooner than what has been observed by these researchers. This proposal fits with recent reports that young children's performance in non-symbolic comparison (Mussolin et al., 2012) and approximate addition (Gilmore et al., 2010) tasks significantly relates to their exact numerical abilities. It is, however, at odds with the predominant failure to find systematic associations between ANS and symbolic arithmetic learning (e.g., Piazza et al., 2010; Soltesz et al., 2010). As mentioned earlier, we propose that the positive findings with the present design might be due to the combination of several methodological parameters (i.e., no number-ratio restriction, high number of repetitions, no participant rejection), which optimize the paradigm's sensitivity to individual differences in ANS acuity. According to this interpretation we should also be able to observe this relation for primary school children (e.g., Holloway and Ansari, 2009; Soltesz et al., 2010; De Smedt and Gilmore, 2011) if sufficiently fine-grained assessment methods are used.

Whereas the symbolic estimation (i.e., "dots to AN") only related to 3rd grade preschooler's early number competence, non-symbolic estimation (i.e., performance in the "dots to dots" task) was related to non-symbolic early number competence in both grades. Positive correlation between performance in the two types of non-symbolic tasks can easily be explained by the common low-level perception processes that underlie task performance and are present from an early age, independently of preschool education and the understanding of the verbal counting system (Brannon and Van de Walle, 2001; Rousselle et al., 2004). In contrast, the fact that symbolic estimation only related to early number competence at the end of Kindergarten indicates that the integration of approximate and exact representations of number symbols is emerging later in development. Combined with the (relatively) late influence of the school's SEI on estimation abilities, it suggests that the innate ANS might initially be impervious to external influences and further supports our hypothesis that newly learned exact symbolic number competences retro-influence ANS precision. Indeed, estimation abilities of 2nd grade preschoolers did not correlate with their early number competences [4] (which depend on formal instruction), nor were they influenced by contextual factors (such as the school's SEI). From 2nd to 3rd Kindergarten grade estimation abilities then globally improved and achieved a higher maturational level, at which school SEI and pre-mathematical instruction interacted with estimation performances. This suggests that early ANS abilities are relatively insensitive to external

factors (such as education and SEI) while maturing up to a certain developmental stage, here 3rd Kindergarten grade. Only once this maturational stage has been reached, the ANS is then systematically affected by the mastery of number symbols and SEI, amongst others. These developmental outcomes are similar to the results observed in adult Munduruku, since data on this Amazonian indigene group show that Munduruku with a certain level of symbolic number knowledge have a more refined ANS than their completely un-educated/instructed peers (Pica et al., 2004). They are also in line with a recent study showing that approximate number skills are less precise in western adults who did not received a formal math education that in their math-educated peers (Nys et al., 2013). Taken together, they support the idea that the ANS serves as cognitive scaffold for the development of the exact symbolic system, especially if future studies could highlight a developmental switch from over-toward underestimation, which is expected to occur when the exact symbolic number system has been acquired through instruction.

Finally it is worth noting that our study also provides insights into the directions that should be taken for developing optimal educational strategies. Taken at face-value the present results indeed suggest that early numeracy interventions should focus on developing good exact symbolic knowledge and then reinforce the link between the innate number sense and those learned symbolic skills.

CONCLUSION

Our study addressed the relationship between approximate number processing and the exact number knowledge in Kindergarten children coming from different socio-economic environments. By investigating for the first time preschooler's abilities to estimate numerosities which largely exceed the numerical values they master exactly we found surprising estimation abilities at a very young age, i.e., from 5 years on.

Compared to their 2nd grade peers, children attending the 3rd grade of Kindergarten produced more accurate symbolic and non-symbolic estimates. In this group, which was close to entering primary school (and formal math education) we also observed a robust relationship between exact symbolic knowledge and ANS acuity. Moreover estimation abilities of 3rd grade Kindergarten children were influenced by the socio-economic context. This relatively late effect on ANS contrasted with the observation that SEI already influenced children's exact early number competence in 2nd Kindergarten grade.

Using a free estimation approach allowed us to disclose a link between the ANS and early number competences, which seems more difficult to highlight with other paradigms. Accordingly, we propose that this method is a very promising tool to obtain further direct insights into the characteristics and development of children's ANS representations.

---

[4]Except for the correlation between performances in non-symbolic tasks, due to shared perceptual processes.

REFERENCES

Barth, H., La Mont, K., Lipton, J. S., Dehaene, S., Kanwisher, N., and Spelke, E. S. (2006). Non-symbolic arithmetic in adults and young children. *Cognition* 98, 199–222. doi: 10.1016/j.cognition.2004.09.011

Barth, H., La Mont, K., Lipton, J. S., and Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14116–14121. doi: 10.1073/pnas.0505512102

Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 42, 189–201. doi: 10.1037/0012-1649.41.6.189

Brannon, E. M., and Van de Walle, G. A. (2001). The development of ordinal numerical competence in young children. *Cogn. Psychol.* 43, 53–81. doi: 10.1006/cogp.2001.0756

Cantlon, J. F. (2012). Math, monkeys, and the developing brain. *Proc. Natl. Acad. Sci. U.S.A.* 109(Suppl. 1), 10725–10732. doi: 10.1073/pnas.1201893109

Castronovo, J., and Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PLoS ONE* 7:e33832. doi: 10.1371/journal.pone.0033832

Castronovo, J., and Seron, X. (2007). Numerical estimation in blind subjects: evidence of the impact of blindness and its following experience. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1089–1106. doi: 10.1037/0096-1523.33.5.1089

Chillier, L. (2002). "La ligne numérique et les codages du nombre chez l'enfant," in *Le Développement Des Activités Numériques Chez L'enfant*, eds J. Bideaud and H. Lehalle (Paris: Hermes, Lavoisier), 129–150.

Cordes, S., Gelman, R., Gallistel, C. R., and Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychon. Bull. Rev.* 8, 698–707. doi: 10.3758/BF03196206

Crollen, V., Castronovo, J., and Seron, X. (2011). Under- and over-estimation. *Exp. Psychol.* 58, 39–49. doi: 10.1027/1618-3169/a000064

Crollen, V., and Seron, X. (2012). Over-estimation in numerosity estimation tasks: more than an attentional bias. *Acta Psychol.* 140, 246–251. doi: 10.1016/j.actpsy.2012.05.003

De Smedt, B., and Gilmore, C. K. (2011). Defective number module or impaired access. numerical magnitude processing in first graders with mathematical difficulties. *J. Exp. Child Psychol.* 108, 278–292. doi: 10.1016/j.jecp.2010.09.003

De Smedt, B., Verschaffel, L., and Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *J. Exp. Child Psychol.* 103, 469–479. doi: 10.1016/j.jecp.2009.01.010

de Villers, J., and Desagher, C. (2011). "L'indice socio-economique des écoles. Comment ça marche? A quoi ça sert?," in *Les analyses de la FAPEO*, Available online at: http://www.fapeo.be

Dossey, J. A., Mullis, I. V. S., Lindquist, M. M., and Chambers, D. L. (1988). "The mathematics report card: are we measuring up?," in *National Assessment of Educational Progress Report*, ed E. T. Service (Princeton, NJ).

Gilmore, C. K., McCarthy, S. E., and Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature* 447, U589–U591. doi: 10.1038/nature05850

Gilmore, C. K., McCarthy, S. E., and Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition* 115, 394–406. doi: 10.1016/j.cognition.2010.02.002

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., and Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11116–11120. doi: 10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M. M., and Feigenson, L. (2008). Individual differences in nonverbal number acuity correlate with maths achievement. *Nature* 455, U665–U668. doi: 10.1038/nature07246

Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children's mathematics achievement. *J. Exp. Child Psychol.* 103, 17–29. doi: 10.1016/j.jecp.2008.04.001

Huntley-Fenner, G. (2001). Children's understanding of number is similar to adults' and rats': numerical estimation by 5-7-year-olds. *Cognition* 78, B27–B40.

Izard, V., and Dehaene, S. (2008). Calibrating the mental number line. *Cognition* 106, 1221–1247. doi: 10.1016/j.cognition.2007.06.004

Izard, V., Sann, C., Spelke, E. S., and Streri, A. (2009). Newborn infants perceive abstract numbers. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10382–10385. doi: 10.1073/pnas.0812142106

Jordan, N. C., Kaplan, D., Oláh, L. N., and Locuniak, M. N. (2006). Number sense growth in Kindergarten: a longitudinal investigation of children at risk for mathematics difficulties. *Child Dev.* 77, 153–175.

Lass, U., Lüer, G., Ulrich, M., and Werner, S. (1993). "Access to analog representations in memory for visually perceived forms: the facilitating effect of declarative knowledge," in *The Cognitive Psychology of Knowledge*, eds G. Strube and K. F. Wender (Amsterdam: The German Wissenspsychologie project), 75–96.

Libertus, M. E., Feigenson, L., and Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Dev. Sci.* 14, 1292–1300. doi: 10.1111/j.1467-7687.2011.01080.x

Mejias, S., Grégoire, J., and Noël, M.-P. (2012a). Numerical estimation in adults with and without developmental dyscalculia. *Learn. Individ. Dif.* 22, 164–170. doi: 10.1016/j.lindif.2011.09.013

Mejias, S., Mussolin, C., Rousselle, L., Gregoire, J., and Noël, M.-P. (2012b). Numerical and nonnumercial estimation in children with and without mathematical learning disabilities. *Child Neuropsychol.* 18, 550-575. doi: 10.1080/09297049.2011.625355

Mundy, E., and Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *J. Exp. Child Psychol.* 103, 490–502. doi: 10.1016/j.jecp.2009.02.003

Mussolin, C., Nys, J., Leybaert, J., and Content, A. (2012). Relationships between approximate number system acuity and early symbolic number abilities. *Trends Neurosci. Educ.* 1, 21–31. doi: 10.1016/j.tine.2012.09.003

Nys, J., Ventura, P., Fernandes, T., Querido, L., Leybaert, J., and Content, A. (2013). Does math education modify the approximate number system. a comparison of schooled and unschooled adults. *Trends Neurosci. Educ.* 2, 13–22. doi: 10.1016/j.tine.2013.01.001

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012

Pica, P., Lemer, C., Izard, V., and Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306, 499–503. doi: 10.1126/science.1102085

Platt, J. R., and Johnson, D. M. (1971). Localization of position within a homogeneous behavior chain: effects of error contingencies. *Learn. Motiv.* 2, 386–414. doi: 10.1016/0023-9690(71)90020-8

Ramani, G., and Siegler, R. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Dev.* 79, 375–394. doi: 10.1111/j.1467-8624.2007.01131.x

Rivera-Batiz, F. L. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *J. Hum. Res.* 27, 313–328. doi: 10.2307/145737

Rousselle, L., and Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition* 102, 361–395. doi: 10.1016/j.cognition.2006.01.005

Rousselle, L., and Noël, M.-P. (2008). The development of automatic numerosity processing in preschoolers: evidence for numerosity-perceptual interference. *Dev. Psychol.* 44, 544–560. doi: 10.1037/0012-1649.44.2.544

Rousselle, L., Palmers, E., and Noël, M.-P. (2004). Magnitude comparison in preschoolers: what counts. influence of perceptual variables. *J. Exp. Child Psychol.* 87, 57–84. doi: 10.1016/j.jecp.2003.10.005

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number, Äìspace mappings: what underlies mathematics achievement. *J. Exp. Child Psychol.* 114, 418–431. doi: 10.1016/j.jecp.2012.10.012

Schneider, W., Eschmann, A., and Zuccolotto, A. (2002). *Version 1.2*. Pittsburgh, PA: Psychology Software Tools Inc.

Soltesz, F., Szucs, D., and Szucs, L. (2010). Relationships between magnitude representation, counting and memory in 4- to 7-year-old children: a developmental study. *Behav. Brain Funct.* 6, 13. doi: 10.1186/1744-9081-6-13

Starkey, P., and Cooper, R. G. (1980). Perception of numbers by human infants. *Science* 210, 1033–1035. doi: 10.1126/science.7434014

Szucs, D., Nobes, A., Devine, A., Gabriel, F. C., and Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. [Original Research]. *Front. Psychol.* 4:444. doi: 10.3389/fpsyg.2013.00444

Whalen, J., Gallistel, C. R., and Gelman, R. (1999). Non verbal counting in humans: the psychophysics of number representation. *Psychol. Sci.*

10, 130–137. doi: 10.1111/1467-9280.00120

Xu, F., and Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition* 74, B1–B11. doi: 10.1016/S0010-0277(99)00066-9

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Developmental changes in the association between approximate number representations and addition skills in elementary school children

Jan Lonnemann[1,2]*, Janosch Linkersdörfer[1,2], Marcus Hasselhorn[1,2,3] and Sven Lindberg[1,2]

[1] Department of Education and Human Development, German Institute for International Educational Research (DIPF), Frankfurt am Main, Germany
[2] Center for Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt am Main, Germany
[3] Department of Educational Psychology, Institute for Psychology, Goethe-University, Frankfurt am Main, Germany

The approximate number system (ANS) is assumingly related to mathematical learning but evidence supporting this assumption is mixed. The inconsistent findings might be attributed to the fact that different measures have been used to assess the ANS and mathematical skills. Moreover, associations between the performance on a measure of the ANS and mathematical skills may be discontinuous, i.e., stronger for children with lower math scores than for children with higher math scores, and may change with age. The aim of the present study was to examine the development of the ANS and arithmetic skills in elementary school children and to investigate how the relationship between the ANS and arithmetic skills develops. Individual markers of children's ANS (internal Weber fractions and mean reaction times in a non-symbolic numerical comparison task) and addition skills were assessed in their first year of school and 1 year later. Children showed improvements in addition performance and in the internal Weber fractions, whereas mean reaction times in the non-symbolic numerical comparison task did not change significantly. While children's addition performance was associated with the internal Weber fractions in the first year, it was associated with mean reaction times in the non-symbolic numerical comparison task in the second year. These associations were not found to be discontinuous and could not be explained by individual differences in reasoning, processing speed, or inhibitory control. The present study extends previous findings by demonstrating that addition performance is associated with different markers of the ANS in the course of development.

Keywords: approximate number system, non-symbolic numerical comparison, arithmetic, development, elementary school

INTRODUCTION

Approximate number representations enable us to discriminate between sets of different numerical quantities, a crucial ability for everyday life. Similar to our performance in discriminating physical dimensions like line length or pitch (e.g., Henmon, 1906), comparing numerical magnitudes is ratio-dependent. We are faster and more accurate in comparing dot arrays with respect to their quantity the smaller the ratio between them is (when dividing the smaller numerosity by the larger one; e.g., van Oeffelen and Vos, 1982). The ability to discriminate between different numerical quantities is present early in life and undergoes a progressive refinement throughout development: in their first hours of life, infants seem to be sensitive to a ratio of 1:3 (Izard et al., 2009) and the precision increases to a ratio of about 9:10 or 10:11 at the age of 20 years (Halberda and Feigenson, 2008). Besides, animals such as monkeys or fish also seem to be able to represent and compare numerical quantities showing similar performance patterns as human adults (Cantlon and Brannon, 2006; Agrillo et al., 2012). This suggests the existence of an evolutionary ancient, innate system, the approximate number system [ANS; see Piazza (2010), for an overview].

It is assumed that the ANS encodes numerosities as analog magnitudes that can be modeled as overlapping Gaussian distributions of activations on a logarithmically compressed internal continuum (Dehaene et al., 2003; Piazza et al., 2004; see e.g., Gallistel and Gelman, 1992, for a different view). Due to the logarithmic compression, overlap between numerosities increases with magnitude, which concurs with a decrease in discriminability. An established measure of the ability to discriminate numerosities and therefore of the precision of the internal representation is the so-called "internal Weber fraction," which reflects the width of the Gaussian distributions. The Weber fraction measures the smallest numerical difference that can be reliably detected, and equals the difference between the two numerosities divided by the smaller numerosity [e.g., 1:3, $(3 - 1)/1 = 2$; 7:8, $(8 - 7)/7 = 0.14$].

ANS precision does not only vary across development but also between individuals of the same age and it has been hypothesized that these inter-individual differences are linked to mathematical skills (e.g., Halberda et al., 2008). However, evidence supporting this proposal is inconsistent. While a number of studies showed that inter-individual differences in performance on a measure

of the ANS are related to concurrent and future mathematics achievement, other studies failed to find such relationships (see De Smedt et al., 2013, for an overview). These divergences might be attributed to the fact that different measures have been used to assess the ANS. The range of applied measures includes the internal Weber fractions, mean error rates (ER), mean or median reactions times (RT) as well as distance or ratio effects calculated on the basis of ER or RT in magnitude comparison tasks. According to De Smedt et al. (2013), however, it is not easy to distinguish studies that *have* from those studies that *have not* found significant relationships between the performance on a measure of the ANS and mathematical skills on the basis of the ANS measure employed. In the case of examining children, those studies that used the internal Weber fractions as dependent measure predominantly detected associations with mathematical skills (see **Table 1** in De Smedt et al., 2013). Recent studies, however, suggest that these associations are limited to trials of non-symbolic numerical comparison tasks in which the size of the area occupied by the stimuli conflicts with the number of elements (i.e., more numerous stimuli occupy a smaller area). Hence, it was inferred that the association represents an artifact of the inhibitory control demands of theses trials and it could be demonstrated that the correlation became non-significant when controlling for inhibitory control (Gilmore et al., 2013; Wagner Fuhs and McNeil, 2013). Besides different measures assessing the ANS, another possible explanation for the inconsistent findings may be the measure used to assess mathematical skills. Typically, standardized or curriculum measures of mathematics achievement have been employed assessing a range of different mathematical competences. De Smedt et al. (2013) argue that the ANS might, however, be more important for some aspects of mathematical competencies than others, and therefore, associations with specific measures of mathematical performance need to be explored. As indicated by a recent study, associations may also be discontinuous, i.e., stronger for children with lower math scores than for children with higher math scores (see Bonny and Lourenco, 2013). Furthermore, the associations may change with age. As, however, most of the studies looking for associations between the performance on a measure of the ANS and

mathematical skills are cross-sectional, potential intra-individual changes have not been revealed. To our knowledge, there is only one longitudinal study examining the development of the association between the ANS and mathematical skills (see Libertus et al., 2013). Individual markers of preschool children's ANS (i.e., the internal Weber fractions and mean RT in a non-symbolic numerical comparison task) and mathematical skills (i.e., counting, comparison of spoken number words, reading Arabic numerals, as well as mental and written calculation) were assessed twice, with a 6-month delay, and improvements in all measures could be detected. Moreover, associations between the ANS and mathematical skills were found at both time points and the ANS was found to predict math ability even when controlling for individual differences in math ability at the initial testing.

The aim of the present study was to expand this evidence by investigating elementary school children. Indeed, evidence on the intra-individual development of the ANS in elementary school children is missing and it is still unclear how the relationship between the ANS and mathematical skills develops in these children. In order to address these issues, we assessed individual markers of children's ANS and mathematical skills in their first year of school and 1 year later. We used the internal Weber fractions and mean RT in a non-symbolic numerical comparison task as markers of the ANS and decided to concentrate specifically on children's performance in addition tasks as addition represents an essential mathematical skill children learn in the first years of elementary school. To assure that possible associations could not be explained by individual differences in more general performance factors, reasoning abilities were also assessed. Moreover, a visual detection task was used to rule out that possible associations could solely be ascribed to individual differences in general processing speed, what might be the case for an association between mean RT in the non-symbolic numerical comparison task and addition skills. As recent studies suggest that a relationship between the performance on a measure of the ANS and mathematical skills might be an artifact of inhibitory control demands (Gilmore et al., 2013; Wagner Fuhs and McNeil, 2013), a visual Go/NoGo task was used to assess inhibitory control.

Table 1 | Comparison of first and second-year performance (paired-sample *t*-tests) with respect to the internal Weber fractions (w comparison), reaction times (in ms) in the non-symbolic numerical comparison task (RT comparison), the addition task, reaction times (in ms) in the visual detection task (proc. speed) as well as reaction times (in ms), omission and commission errors (in %) in the inhibitory control task (RT inhib., ER om. inhib., ER com. inhib.).

	First year			Second year			*p* (Two-sided)
	M	*SD*	*SE*	*M*	*SD*	*SE*	
w comparison	0.35	0.14	0.02	0.25	0.10	0.01	$p < 0.001$
RT comparison	1362	328	40	1321	261	32	$p = 0.25$
Addition	51	22	2.7	70	21	2.5	$p < 0.001$
proc. speed[a]	510	96	12	507	98	12	$p = 0.79$
RT inhib.[a]	730	57	7	697	62	8	$p < 0.001$
ER om. inhib.[a]	7.9	7.0	0.86	5.9	5.6	0.68	$p < 0.05$
ER com. inhib.[a]	5.2	5.2	0.64	5.0	4.4	0.54	$p = 0.75$

n = 67; [a]n = 66.

MATERIALS AND METHODS

PARTICIPANTS

Sixty-seven children (35 girls and 32 boys) completed all tasks at both measurement time points. At the first measurement time point, all children were first-graders (mean age: 87 months; range: 79–94 months) and at the second measurement time point second-graders (mean age: 98 months; range: 90–106 months). The average delay between individual measurement time points was 355 days. Written and informed consent was obtained from all parents involved.

MATERIALS

All tasks were carried out individually. Apart from the measure of reasoning abilities which was administered only at the second measurement time point, all tasks were carried out at both measurement time points.

NON-SYMBOLIC NUMERICAL COMPARISON

Sets of black dots were presented in two white circles on the left and the right hand side of the screen of a 14-inch notebook running Presentation® software (Neurobehavioral Systems, Inc.). From a viewing distance of about 60 cm, each of the white circles had a visual angle of 7.82° (82 mm) and the black dots ranged between 0.10 and 0.14° (1–1.5 mm). On each trial, one of the white circles contained 32 dots (reference numerosity) and the other one 20, 23, 26, 29, 35, 38, 41, or 44 dots (deviants). Each of these eight comparison pairs appeared eight times, four times with the reference numerosity on the left and four times on the right hand side. Every single comparison pair had a unique configuration of dots. In half of the 8 trials per comparison pair, the size of the area occupied by the dots in each circle was held constant (luminance-controlled trials), while in the other half, individual dot size in each circle was held constant (size-controlled trials). Children were asked to indicate without using counting strategies, the side of the larger numerical magnitude by answering with the left index finger when it was larger on the left hand side and by using the right index finger when it was larger on the right hand side. Responses were given by pressing the left and right CTRL-buttons of the notebook's keyboard. RT and ER were recorded, and the instruction stressed both speed and accuracy. The order of trials was pseudo-randomized so that there were no consecutive identical comparison pairs. The experiment started with eight warm-up trials to familiarize children with the task (data not recorded), followed by a total of 64 experimental trials (8 comparison pairs × 2 perceptual control conditions × 4 repetitions). A trial started with the presentation of a black screen for 700 ms. After the black screen had vanished, the target appeared until a response was given, but only up to a maximum duration of 4000 ms. No feedback regarding the correctness of responses was provided. Mean RT and internal Weber fractions were used as individual markers of the ANS (see Halberda et al., 2012; Libertus et al., 2013). The internal Weber fractions were calculated based on ER for eight different ratios (20/32, 23/32, 26/32, 29/32, 35/32, 38/32, 41/32, and 44/32) following the methods described in the Supplemental Data from Piazza et al. (2004). The calculation was based on the formula $y = 0.5*(1 + \mathrm{erf}(\log(x) / (\mathrm{sqrt}(2)* w)))$, where y is the probability of responding "larger" and x are the different ratios.

ADDITION

We used a subtest of the standardized German scholastic achievement test for mathematics (DIRG; Grube et al., 2010) that includes 110 simple addition problems in which two single-digit numbers (excluding 0 and 1) have to be added. Solutions range from 5 to 10 and ties (e.g., $4 + 4$) are not included. The 110 addition problems consist of 24 different problems presented in pseudo-randomized order ensuring that neither identical nor commutated problems follow each other directly. The repetition rate of the different tasks varies (some problems are only presented three times, while others are presented up to six times). The problems were presented in written form on four different pages. Children were asked to write down as many solutions as possible in 4 min adhering to the order of the pages. Addition performance was calculated as the number of correctly answered problems. Total scores ranging from 0 to 110 are reported for each child.

REASONING

Raven's Colored Progressive Matrices (CPM; Bulheller and Häcker, 2002) were used to assess inductive reasoning. The CPM is an untimed power test consisting of 36 colored diagrammatic puzzles, each with a missing part which has to be identified from a choice of six. Total scores ranging from 0 to 36 are reported for each child.

PROCESSING SPEED

A visual detection task was used to assess individual processing speed. Children were instructed to press the space bar of the notebook's keyboard as fast as possible whenever an "×" appeared in the center of the screen. The target appeared until a response was given, but only up to a maximum duration of 3000 ms. The task comprised 30 experimental trials with varying inter-trial intervals (2000, 3500, 5000, 6500, or 8000 ms). Mean RT is reported for each child.

INHIBITORY CONTROL

A visual Go/NoGo task was used to assess inhibitory control. Children were instructed to press the space bar of the notebook's keyboard as fast as possible whenever an "×" appeared in the center of the screen (Go-trials) and to inhibit responses whenever an "+" appeared in the center of the screen (NoGo-trials). The target appeared until a response was given, but only up to a maximum duration of 3000 ms. The task comprised 40 experimental trials (20 Go-trials and 20 NoGo-trials) with varying inter-trial intervals (2000, 3500, 5000, 6500, or 8000 ms). The order of trials was pseudo-randomized so that there were no more than three consecutive identical trials. Mean RT, mean commission ER (button presses in NoGo-trials), and mean omission ER (no button presses in Go-trials) are reported for each child.

RESULTS

Only trials with correct responses were used for computing mean RT in the non-symbolic numerical comparison task, in the visual detection task, and in the inhibitory control task. Trials in which

the response was either given too late (after 4000 ms in the non-symbolic numerical comparison task and after 3000 ms in the visual detection task as well as in the inhibitory control task) or not at all were classified as errors. Responses below 200 ms were excluded from further analysis. This resulted in 0.06% of response exclusions in the non-symbolic numerical comparison task, in 0.22% of response exclusions in the visual detection task, and in 0.06% of response exclusions in the inhibitory control task. Mean ER in the visual detection task was low (first year: 1.5%; second year: 2%) and not further analyzed. Pearson correlation coefficients were computed for the observed variables. To assure that possible correlations between individual markers of the ANS and addition performance within the respective years could not be explained by individual differences in more general performance factors or in inhibitory control, partial correlations were computed controlling for reasoning abilities which were only assessed in the second year, as well as for processing speed and inhibitory control (mean RT, omission and commission ER) of the respective year. Correlations examining the predictive value of markers of the ANS for addition performance of the second year were controlled for reasoning abilities as well as for processing speed, inhibitory control, and addition performance of the first year. Moreover, correlations between the internal Weber fractions of the second year and addition performance of the first year were controlled for the internal Weber fractions of the first year as well as for reasoning abilities, processing speed, and inhibitory control of the second year and correlations between mean RT in the non-symbolic numerical comparison task of the second year and addition performance of the first year were controlled for mean RT in the non-symbolic numerical comparison task of first year as well as for reasoning abilities, processing speed, and inhibitory control of the second year. In the partial correlation analyses including performance in the visual detection and the inhibitory control task at the second measurement time point, one child had to be excluded because of failing to complete these tasks. To test for the possibility that correlations between individual markers of the ANS and addition skills may be stronger for children with lower addition scores than for children with higher addition scores (see Bonny and Lourenco, 2013), we conducted segmented regression analyses using the software SegReg (Oosterbaan, 2011). These analyses allowed us to look for possible breakpoints in the addition performance where the relation with the markers of the ANS changes abruptly. We looked for models with two lines with different slopes or models with a sloping segment followed by a horizontal line. Evidence for a breakpoint would be reflected in greater explained variance compared with a single linear model.

Significant improvements were observed for the internal Weber fractions, the addition performance, mean RT, and mean omission ER in the inhibitory control task but not for mean commission ER in the inhibitory control task, mean RT in the visual detection task, and mean RT in the non-symbolic numerical comparison task (see **Table 1**).

In the non-symbolic numerical comparison task, ER increased as the ratio between the two to-be-compared numerosities increased: significant linear trends for deviants smaller than the reference [20 vs. 23 vs. 26 vs. 29; first year: $F_{(1, 66)} = 183.14$;

$p < 0.001$; second year: $F_{(1, 66)} = 230.97$; $p < 0.001$] and for deviants larger than the reference [35 vs. 38 vs. 41 vs. 44; first year: $F_{(1, 66)} = 38.96$; $p < 0.001$; second year: $F_{(1, 66)} = 36.61$; $p < 0.001$] were found in both years (see **Figure 1**). In order to look for differences between luminance-controlled and size-controlled trials, we computed the internal Weber fractions and mean RT for both conditions separately. Because of a very low fitting parameter ($R^2 < 0.2$) of the procedure to calculate the internal Weber fractions, nine children had to be excluded in the first year and 11 children in the second year. As a consequence, we used mean ER as a proxy for the internal Weber fractions (see Mazzocco et al., 2011, for a similar approach; mean ER and the internal Weber fractions for all trials were highly correlated in both years: first year: $r = 0.95$; $p < 0.001$ [two-sided]; second year: $r = 0.94$; $p < 0.001$ [two-sided]) which allowed us to compare the performance in luminance-controlled and size-controlled trials in all participants. Considering mean ER, a significant difference between luminance-controlled and size-controlled trials was found in the second year (mean ER luminance-controlled = 21% vs. mean ER size-controlled = 16%, $p < 0.001$ [two-sided]) and a trend toward a significant difference in the first year (mean ER luminance-controlled = 25% vs. mean ER size-controlled = 23%, $p = 0.11$ [two-sided]). However, in the first and in the second year, ER increased as the ratio between the two to-be-compared numerosities increased for both luminance-controlled and size-controlled trials (significant linear trends for deviants smaller than the reference [20 vs. 23 vs. 26 vs. 29; luminance-controlled—first year: $F_{(1, 66)} = 97.10$; $p < 0.001$; second year: $F_{(1, 66)} = 117.39$; $p < 0.001$; size-controlled—first year: $F_{(1, 66)} = 98.72$; $p < 0.001$; second year: $F_{(1, 66)} = 230.97$; $p < 0.001$] and for deviants larger than the reference [35 vs. 38 vs. 41 vs. 44; luminance-controlled—first year: $F_{(1, 66)} = 21.10$; $p < 0.001$; second year: $F_{(1, 66)} = 11.76$; $p = 0.001$; size-controlled—first year: $F_{(1, 66)} = 22.62$; $p < 0.001$; second year: $F_{(1, 66)} = 37.93$; $p < 0.001$]). Mean RT in luminance-controlled and size-controlled trials did not significantly differ in both years (first year: mean RT luminance-controlled = 1337 ms vs. mean RT size-controlled = 1324 ms, $p = 0.43$ [two-sided]; second year: mean RT luminance-controlled = 1320 ms vs. mean RT size-controlled = 1304, $p = 0.40$ [two-sided]).

Correlation coefficients for the observed variables are shown in **Table 2**. Significant positive correlations were found between the variables that had been assessed twice, with the exception

FIGURE 1 | Mean error rates (ER) and standard errors in % in the non-symbolic numerical comparison task, separately for the different measurement points (first year, second year) as a function of the different deviants (reference numerosity = 32).
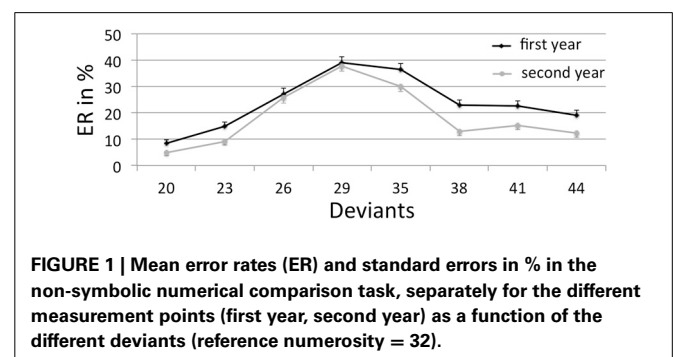
Table 2 | Bivariate (below the diagonal) correlation coefficients for all observed variables and partial (above the diagonal) correlation coefficients between individual markers of the ANS (internal Weber fractions [w comparison] and mean reaction times [RT comparison] in the non-symbolic numerical comparison task) and addition performance within and across the respective years (processing speed [proc. speed]; reaction times, omission, and commission errors in the inhibitory control task [RT inhib., ER om. inhib., ER com. inhib.]).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Addition first year	–		−0.39*	0.18[a]	−0.01	−0.33[a]**									
2. Addition second year	0.75**	–	−0.02	0.08[a]	−0.10	−0.30[a]*									
3. w comparison first year	−0.39*	−0.27*	–												
4. w comparison second year	0.03	0.03	0.15	–											
5. RT comparison first year	−0.15	−0.17	−0.10	0.16	–										
6. RT comparison second year	−0.38**	−0.31*	0.20	0.03	0.54**	–									
7. proc. Speed first year	−0.32**	−0.25*	0.15	0.08	0.26*	0.27*	–								
8. proc. speed second year	−0.28*	−0.18	0.19	0.27*	0.08	0.29*	0.26*	–							
9. RT inhib. first year	−0.30*	−0.20	0.15	0.12	0.31*	0.35**	0.42**	0.23	–						
10. RT inhib. second year	−0.26*	−0.19	0.23	0.13	0.30*	0.47**	0.17	0.56**	0.36**	–					
11. ER com. inhib. first year	0.03	0.01	0.07	0.13	−0.33*	−0.30*	−0.15	0.07	−0.31*	−0.16	–				
12. ER com. inhib. second year	0.01	0.07	0.01	0.31*	−0.16	−0.18	0.19	−0.07	0.01	−0.36**	0.26*	–			
13. ER om. inhib. first year	−0.31*	−0.21	0.28*	0.20	0.38**	0.35**	0.61**	0.23	0.49**	0.30*	−0.04	0.04	–		
14. ER om. inhib. second year	−0.19	−0.07	0.33**	0.31*	0.18	0.27*	0.31*	0.51**	0.29*	0.28*	0.06	0.11	0.48*	–	
15. Reasoning second year	0.40**	−0.44**	−0.01	−0.01	−0.12	−0.09	−0.22	−0.31*	−0.26*	−0.20	−0.14	0.03	−0.25*	−0.18	–

$n = 67$; [a]$n = 66$; *$p < 0.05$ (two-sided); **$p < 0.01$ (two-sided).

of the internal Weber fractions. Significant correlations in the bivariate and in the partial correlation analyses were found between the internal Weber fractions and the addition performance in the first year (similarly, mean ER in the non-symbolic numerical comparison task and addition performance correlated significantly—bivariate: $r = −0.38$; $p = 0.001$ [two-sided]; partial: $r = −0.36$; $p < 0.01$ [two-sided]) as well as between mean RT in the non-symbolic numerical comparison task and the addition performance in the second year (see **Figure 2**). There was no trade-off between mean RT and mean ER in the non-symbolic numerical comparison task (first year: $r = −0.11$; $p = 0.37$ [two-sided]; second year: $r = −0.01$; $p = 0.94$ [two-sided]). Significant correlations in the bivariate and in the partial correlation analyses across the 2 years were only found between addition performance in the first year and mean RT in the non-symbolic numerical comparison task in the second year. A significant correlation between the internal Weber fractions of the first year and addition performance of the second year was only found in the bivariate but not in the partial correlation analyses.

Regarding the correlation between the internal Weber fractions and the addition performance in the first year, a model with a breakpoint in the addition performance and two lines with different slopes did not fit the data but a model with a breakpoint in the addition performance and a sloping segment followed by a horizontal line did [$F_{(3, 63)} = 3.03$, $p < 0.05$]. This model did, however, explain less variance than a model without a breakpoint in the addition performance [$F_{(1, 65)} = 11.99$, $p = 0.001$]. For the correlation between mean RT in the non-symbolic numerical comparison task and the addition performance in the second year, neither a model with two lines with different slopes nor a model with a sloping segment followed by a horizontal line fit the data.
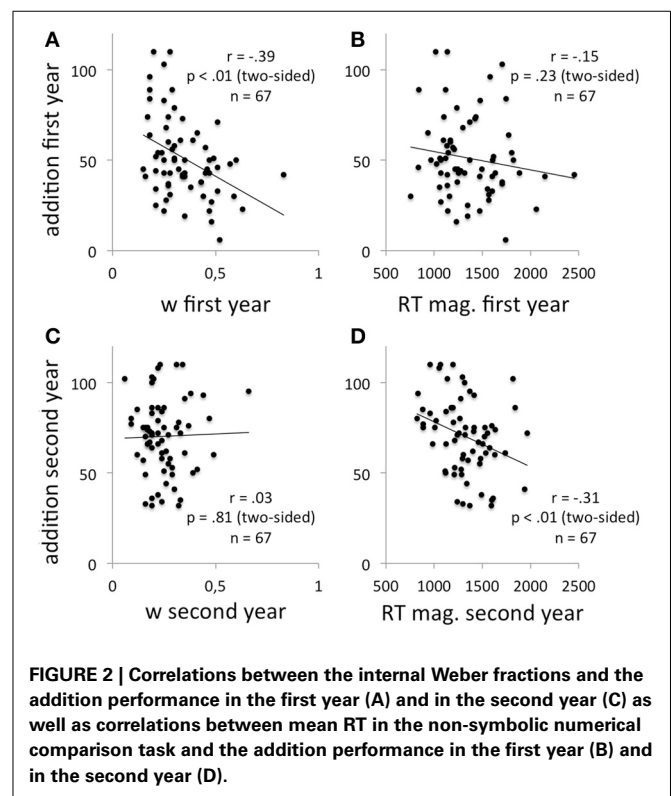
FIGURE 2 | Correlations between the internal Weber fractions and the addition performance in the first year (A) and in the second year (C) as well as correlations between mean RT in the non-symbolic numerical comparison task and the addition performance in the first year (B) and in the second year (D).

As a significant difference between ER in luminance-controlled and size-controlled trials in the second year and a trend toward a significant difference in the first year was observed, we computed correlations with the addition performance separately for both conditions. Significant correlations with the addition performance in the first year could be found for

size-controlled ER (bivariate: $r = -0.41$; $p = 0.001$ [two-sided]; partial: $r = -0.39$; $p < 0.01$ [two-sided]) and trends toward significant associations for luminance-controlled ER (bivariate: $r = -0.21$; $p = 0.09$ [two-sided]; partial: $r = -0.20$; $p = 0.13$ [two-sided]). Comparing the correlation coefficients for the luminance-controlled and the size-controlled ER (Hotelling–Williams test; see Steiger, 1980) did not reveal any significant differences (bivariate: $r = -0.41$ vs. $r = -0.21$; $p = 0.14$ [two-sided]; partial: $r = -0.39$ vs. $r = -0.20$; $p = 0.16$ [two-sided]). In the second year, no significant correlations were found (bivariate—ER luminance-controlled: $r = -0.05$; $p = 0.66$ [two-sided], ER size-controlled: $r = 0.02$; $p = 0.90$ [two-sided]; partial—ER luminance-controlled: $r = 0.20$; $p = 0.12$ [two-sided], ER size-controlled: $r = 0.03$; $p = 0.82$ [two-sided]). Comparing the correlation coefficients for the luminance-controlled and the size-controlled ER again revealed no significant differences (bivariate: $r = -0.05$ vs. $r = 0.02$; $p = 0.65$ [two-sided]; partial: $r = 0.20$ vs. $r = 0.03$; $p = 0.28$ [two-sided]).

DISCUSSION

In the present study, the development of the ANS and addition skills was examined in children in their first 2 years of elementary school. Significant improvements in addition performance and in the internal Weber fractions were found, while mean RT in the non-symbolic numerical comparison task remained unchanged. The developmental change in the internal Weber fractions (from 0.35 to 0.25) is in line with previous findings from cross-sectional studies (see Piazza, 2010). The internal Weber fractions were associated with children's addition performance at the first measurement time point. This association was not found to be non-linear (e.g., stronger for children with lower addition scores than for children with higher addition scores) and it could not be explained by individual differences in reasoning, processing speed, and inhibitory control. At the second measurement time point, however, no association was found between the same measures. Likewise, the internal Weber fractions of the first year were not correlated with the internal Weber fractions of the second year. This might be due to the fact that mean ER in the non-symbolic numerical comparison task, used as a proxy for the internal Weber fractions, was significantly higher in luminance-controlled trials than in size-controlled trials in the second year, whereas no significant difference was found in the first year. The difference between luminance-controlled and size-controlled trials detected in the second year concurs with previous findings (e.g., Wagner Fuhs and McNeil, 2013) and might be related to the fact that luminance and the number of elements are positively correlated in size-controlled trials and uncorrelated in luminance-controlled trials. The visual characteristics of the stimuli could thus, provide an additional cue to number in size-controlled trials, whereas the visual characteristics of the stimuli in luminance-controlled trials might even be obstructive because controlling for luminance involves that the more numerous arrays have smaller dots. As a significant difference between ER in luminance-controlled and size-controlled trials was found only in the second but not in the first year, the influence of the visual cues might have differed at the two measurement

time points, possibly resulting in the non-significant correlation between the internal Weber fractions of the first and of the second year.

In contrast to the present study, some other studies investigating the association between the performance on a measure of the ANS and mathematical skills incorporated a condition in the non-symbolic numerical comparison task in which luminance and the number of elements was negatively correlated (so-called inverse or incongruent trials), either in addition to a luminance- and a size-control condition (Wagner Fuhs and McNeil, 2013) or instead of a luminance-control condition (see experiment 1 in Gilmore et al., 2013). In both cases, ER differed significantly between the respective conditions and a relationship between performance on a measure of the ANS and mathematical skills was limited to the inverse trials. Hence, it was inferred that this relationship represented an artifact of the inhibitory control demands of the inverse trials and it could be demonstrated that the correlation became non-significant when controlling for inhibitory control. These findings do not correspond to the results of the present study. Indeed, we found that mean ER in the non-symbolic numerical comparison task did not differ significantly between luminance-controlled and size-controlled trials in the first year, and the association between the internal Weber fractions (or mean ER respectively) and children's addition performance detected in the same year was not limited to luminance-controlled trials and did not disappear when controlling for inhibitory control. Previous studies used other tasks to measure inhibitory control (see Gilmore et al., 2013; Wagner Fuhs and McNeil, 2013) and thus, it is possible that associations would have disappeared when using another task. Moreover, an inclusion of inverse trials requiring high levels of inhibitory control in the present study might possibly have provoked significant differences between the respective conditions of the non-symbolic numerical comparison task at the first measurement time point. The absence of such a condition as well as the choice of the inhibitory control task can, however, hardly explain why the association between the internal Weber fractions (or mean ER respectively) and children's addition performance at the first measurement time point was not limited to luminance-controlled trials. We assume that this might be due to different measures of mathematical skills used in the respective studies. Instead of selectively assessing a particular proficiency like addition, Wagner Fuhs and McNeil (2013) as well as Gilmore et al. (2013) used test batteries assessing a range of different skills. The internal Weber fractions (or mean ER respectively) might be specifically related to addition skills in first graders and this relationship does not seem to be an artifact of inhibitory control demands.

While the internal Weber fractions (or mean ER respectively) were found to be related to addition skills at the first measurement time point, children's mean RT in the non-symbolic numerical comparison task was associated with their addition performance at the second measurement time point. This association was not stronger for children with lower addition scores than for children with higher addition scores and it could not be explained by individual differences in reasoning, processing speed, and inhibitory control. Moreover, children's mean RT of the first year were significantly correlated with the mean RT of the second year and

mean RT in luminance-controlled trials did not differ from mean RT in size-controlled trials in both years. Consequently, children's performance in solving simple addition tasks seems to be associated with different markers of the ANS in the course of development. This finding contradicts the results of a previous study on preschool children showing that the internal Weber fractions and mean RT in a non-symbolic numerical comparison task were linked to math skills in both of two successive testing sessions (Libertus et al., 2013). According to Halberda et al. (2012), the internal Weber fraction represents an estimate of the ANS's precision while mean RT in a non-symbolic numerical comparison task represents the amount of time it takes individuals to make their decision. Thus, the present findings might indicate that children's addition performance in the first year of school was related to the individual precision of the ANS while addition performance in the second year was related to the individual speed of retrieving approximate number representations. Following the line of argument that the ANS provides semantic representations of numbers (e.g., Dehaene et al., 2003), children might have relied on the ANS during arithmetic problem solving in order to grasp how the magnitudes of the different task solutions (and of the addends) fall in relation to other magnitudes, and whether the solution is appropriate to the task. While not all the children might have grasped this concept with sufficient clarity to adequately process the different addition tasks in the first year, the majority of the children in the second year might have reached the appropriate level of understanding, attributing stronger impact to the speed of retrieval rather than the precision of the representations in the process of solving the addition tasks in the second year. According to this reasoning, the divergent findings by Libertus et al. (2013) might again be attributed to differences in the measures used to assess children's mathematical skills. Libertus et al. (2013) used a test battery involving counting, comparison of spoken number words, reading Arabic numerals, as well as mental and written calculation. Indeed, using the performance of similar tasks such as the comparison of spoken number words as indicator of mathematical skills and the comparison of non-symbolic numerosities as marker of the ANS may increase the chance of detecting a relationship. This may also explain why Libertus et al. (2013) found that individual markers of the ANS predicted mathematical skills at the second measurement time point in preschool, while no reliable evidence for a prediction of arithmetic skills could be detected in the present study. In fact, as arithmetic skills were found to predict mean RT in the non-symbolic numerical comparison task, results of the present study rather point to the reverse direction of influence. Libertus et al. (2013) also reported a similar relationship between mathematical skills at the first measurement time point and the internal Weber fractions at the second measurement time point. Likewise, a recent study revealed that the acquisition of symbolic numbers and arithmetic enhances the precision of the ANS (Piazza et al., 2013). It can thus be assumed that symbolic and non-symbolic numerical thinking enhance one another over the course of development. Looking at developmental trajectories of associations between different markers of the ANS and different mathematical skills might help to better understand what exactly causes the link between the ANS and mathematical performance. In this regard, the present study extends previous findings by demonstrating that the performance in solving simple addition tasks is associated with different markers of the ANS in the course of development.

ACKNOWLEDGMENTS

REFERENCES

Agrillo, C., Piffer, L., Bisazza, A., and Butterworth, B. (2012). Evidence for two numerical systems that are similar in humans and guppies. *PLoS ONE* 7:e31923. doi: 10.1371/journal.pone.0031923

Bonny, J. W., and Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: evidence from the preschool years. *J. Exp. Child Psychol.* 114, 375–388. doi: 10.1016/j.jecp.2012.09.015

Bulheller, S., and Häcker, H. O. (2002). *Coloured Progressive Matrices (CPM). Deutsche Bearbeitung und Normierung nach J. C. Raven.* Frankfurt: Pearson Assessment.

Cantlon, J. F., and Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychol. Sci.* 17, 401–406. doi: 10.1111/j.1467-9280.2006.01719.x

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239

De Smedt, B., Noël, M.-P., Gilmore, C., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing relate to individual differences in children's mathematical skills. A review of evidence from brain and behavior. *Trends Neurosci. Educ.* 2, 48–55. doi: 10.1016/j.tine.2013.06.001

Gallistel, C. R., and Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition* 44, 43–74. doi: 10.1016/0010-0277(92)90050-R

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., et al. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE* 8:e67374. doi: 10.1371/journal.pone.0067374

Grube, D., Weberschock, U., Stuff, M., and Hasselhorn, M. (2010). *Diagnostisches Inventar zu Rechenfertigkeiten im Grundschulalter (DIRG).* Göttingen: Hogrefe.

Halberda, J., and Feigenson, L. (2008). Developmental change in the acuity of the "number sense": the approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Dev. Psychol.* 44, 1457–1465. doi: 10.1037/a0012682

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., and Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11116–11120. doi: 10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with math achievement. *Nature* 455, 665–668. doi: 10.1038/nature07246

Henmon, V. A. C. (1906). The time of perception as a measure of differences in sensation. *Arch. Philos. Psychol. Sci. Methods* 8, 5–75.

Izard, V., Sann, C., Spelke, E. S., and Streri, A. (2009). Newborn infants perceive abstract numbers. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10382–10385. doi: 10.1073/pnas.0812142106

Libertus, M. E., Feigenson, L., and Halberda, J. (2013). Is approximate number precision a stable predictor of math ability. *Learn. Individ. Dif.* 25, 126–133. doi: 10.1016/j.lindif.2013.02.001

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

Oosterbaan, R. J. (2011). *SegReg: Segmented Linear Regression with Breakpoint and Confidence Intervals.* Available online at: http://www.waterlog.info/segreg.htm

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends Cogn. Sci.* 14, 542–551. doi: 10.1016/j.tics.2010.09.008

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human

intraparietal sulcus. *Neuron* 44, 547–555. doi: 10.1016/j.neuron.2004.10.014

Piazza, M., Pica, P., Izard, V., Spelke, E. S., and Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychol. Sci.* 24, 1037–1043. doi: 10.1177/0956797612464057

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychol. Bull.* 87, 245–251. doi: 10.1037/0033-2909.87.2.245

van Oeffelen, M. P., and Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Percept. Psychophys.* 32, 163–170. doi: 10.3758/BF03204275

Wagner Fuhs, M. W., and McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: contributions of inhibitory control. *Dev. Sci.* 16, 136–148. doi: 10.1111/desc.12013

On the interrelation of multiplication and division in secondary school children

Stefan Huber[1]*, Ursula Fischer[1], Korbinian Moeller[1] and Hans-Christoph Nuerk[1,2]

[1] Knowledge Media Research Center, Tuebingen, Germany
[2] Department of Psychology, Eberhard Karls University, Tuebingen, Germany

Multiplication and division are conceptually inversely related: Each division problem can be transformed into as a multiplication problem and vice versa. Recent research has indicated strong developmental parallels between multiplication and division in primary school children. In this study, we were interested in (i) whether these developmental parallels persist into secondary school, (ii) whether similar developmental parallels can be observed for simple and complex problems, (iii) whether skill level modulates this relationship, and (iv) whether the correlations are specific and not driven by general cognitive or arithmetic abilities. Therefore, we assessed performance of 5th and 6th graders attending two secondary school types of the German educational system in simple and complex multiplication as well as division while controlling for non-verbal intelligence, short-term memory, and other arithmetic abilities. Accordingly, we collected data from students differing in skills levels due to either age (5th < 6th grade) or school type (general < intermediate secondary school). We observed moderate to strong bivariate and partial correlations between multiplication and division with correlations being higher for simple tasks but nevertheless reliable for complex tasks. Moreover, the association between simple multiplication and division depended on students' skill levels as reflected by school types, but not by age. Partial correlations were higher for intermediate than for general secondary school children. In sum, these findings emphasize the importance of the inverse relationship between multiplication and division which persists into later developmental stages. However, evidence for skill-related differences in the relationship between multiplication and division was restricted to the differences for school types.

Keywords: simple multiplication, complex multiplication, simple division, complex division, arithmetic, skill level, development

INTRODUCTION

Basic arithmetic operations constitute a milestone of numerical development. Children are introduced to addition, subtraction, multiplication, and division in elementary school and consolidate their arithmetic skills throughout secondary school. With schooling, they become more proficient and use more efficient strategies to solve both simple and complex arithmetic operations.

The focus of the current study is on multiplication and division, two of the four basic operations that are mathematically inversely related. By inversion of the operands, each division problem (e.g., $28 \div 4 = 7$) can be recast as a multiplication problem (e.g., $4 \times 7 = 28$) and vice versa. In this study, we aimed to investigate how these two operations interact in fifth and sixth graders. Moreover, we specifically looked at how this interaction is influenced by skill level as the correlation between multiplication and division may not be the same for all children.

To establish how we expect these operations to interact, we will first give a brief overview of the most common strategies used to solve multiplication and division problems before discussing the role of skill level. While different strategies can lead to the correct result, these strategies are not equally efficient and also depend on the complexity of the problem. We will therefore start by outlining strategies used in simple operations before moving onto complex

problems. Thereby, we will also pinpoint how the operations are interrelated.

Different types of multiplication problems are solved by applying different types of strategies. Simple multiplication problems (i.e., problems with single digit factors) are most often solved via fact retrieval from long term memory (e.g., Dehaene et al., 2003). Lemaire and Siegler (1995) reported that as early as grade 2, students' most common strategy to solve simple multiplication problems was the retrieval of multiplication facts. However, there is also evidence suggesting that not all multiplication problems are solved equally well. Multiplication problems with operands 0 or 1 are mostly solved by applying rules (Cooney et al., 1988). Other common findings for multiplications are the problem size effect, the tie effect and the five effect. The problem size effect describes the fact that reaction times and error rates for multiplication problems increase with increasing operands (Campbell and Graham, 1985). The tie effect reflects the finding that problems with equal operands (e.g., 4×4) are solved faster and with fewer errors than other problems (e.g., Campbell and Graham, 1985; Campbell and Gunter, 2002) Similarly, problems including 5 as one of the operands (e.g., 3×5) are solved faster and with fewer errors compared to other problems (five effect; e.g., Siegler, 1988).

When solving complex multiplication problems such as "7 × 16," several steps of cognitive processing may be necessary (Geary and Widaman, 1987; Hope and Sherrill, 1987; Seitz and Schumann-Hengsteler, 2000): (1) Breaking down the multiplication problem ("7 × 10 = ?", "7 × 6 = ?"); (2) retrieval of first fact ("7 × 10 = 70") and storage of first partial result; (3) retrieval of second fact ("7 × 6 = 42") and storage of second partial result; (4) addition of partial results ("70 + 42 = 112"). Thus, besides working memory capacities (for an overview, see Raghubar et al., 2010) and the ability to add the partial results, complex multiplication problems are solved by resorting to the retrieval of simple multiplication facts.

In contrast to multiplication, retrieval of division facts is not a very common strategy until grade 7. Robinson et al. (2006) explored Canadian children's usage of division strategies from grade 4–7 and found that from grade 4 onward, the reported frequency of the retrieval strategy stayed fixed at about 16%. The most common strategy of fourth graders was the addition strategy (i.e., adding up the divisor until the dividend is reached and counting how many times it was added). For instance, when solving "28 ÷ 4," 4 would be added up 7 times (i.e., 4 + 4 + 4 + 4 + 4 + 4 + 4). However, from grade 5 onward the most frequently reported strategy was the division by mediation strategy. Application of this strategy involves relying on the inverse relationship between multiplication and division to solve simple division problems (e.g., "63 ÷ 7 =?"→"7×? = 63"). The frequency of this strategy increases from 48.8% in grade 5–71.0% in grade 7.

In comparison to research on complex multiplication, research on complex division is even sparser (for an exception, see Hickendorff et al., 2010). Complex division problems usually refer to problems containing either a two-digit divisor (e.g., "90 ÷ 18 =5") or a two-digit quotient (e.g., "108 ÷ 6 = 14"). One way to solve complex division problems with a two-digit quotient is to split them up into multiple simple division problems, also known as the long division algorithm which was first described by Euclid (e.g., Heath, 1956). For instance, a typical strategy to solve "108 ÷ 6" would be: (1) Calculating how often 6 fits in 10 (once); (2) multiplying the result with the divisor (i.e., 1 × 6 = 6); (3) subtracting the result of step 2 from 10 (i.e., 10 − 6); (4) Multiplying the result of step 3 by 10 and adding 8 (i.e., 4 ×10 + 8 = 48). (5) Solving the simple division problem "48 ÷ 6" and storage of the result. (5) Combining the result of (1) and (5). Thus, like in complex multiplication, proficiency in solving simple division problems should support the solution of complex division problems.

To sum up, the most common strategy to solve simple division problems is the division by mediation strategy via retrieval of multiplication facts. Application of this strategy suggests that division and multiplication proficiency are interrelated, which has, however, not yet been examined in children after primary school. Furthermore, complex multiplication problems are usually solved by breaking them down into simpler multiplication problems. Similarly, solving complex division problems involves solving of simple division problems.

Aside from strategy use, the association between multiplication and division may also depend on skill level as suggested by the revised identical elements model (Rickard, 2005). The original identical elements model postulated that arithmetic facts with identical operands and results (i.e., identical elements) recruit a common representation in long-term memory regardless of operand order (Rickard et al., 1994; Rickard and Bourne, 1996). For instance, "3 × 4 = 12" and "4 × 3 = 12" would be stored in the same memory node, because operands differ only in their order [but see Butterworth et al. (2001), and Verguts and Fias (2005), for an alternative model explaining order effects]. However, operations with different elements would recruit independent representations. For instance, "3 × 4 = 12" and its inverse division problem "12 ÷ 3 = 4" would be stored in separate nodes, because operands and results differ. Accordingly, division problems should form a unique representation, independent from the corresponding multiplication problems. However, this assumption turned out to only be true for participants with a high skill level, as indicated in the revision of the identical elements model (Rickard, 2005). For adult participants with low and intermediate skill levels, the revised model suggests that division problems are solved by mediation via multiplication.

Whether the relationship between multiplication and division also depends on skill level in children has not been systematically investigated yet. In fact, we found only one (longitudinal) study in primary school children, which actually provided evidence contradicting this idea: De Brauwer and Fias (2009) measured arithmetic performance in multiplication and division of third graders twice a year in two consecutive school years, and of another group of children twice in second grade. To investigate the relationship between multiplication and division in primary school children, they studied the developmental parallels in performance of multiplication and division tasks as well as problem size, five and tie effects in multiplication and division. From grade 3 onward, De Brauwer and Fias (2009) found robust problem size, five, and tie effects for both operations. As these effects showed a similar time course for both operations, they suggested that their results indicate strong parallels between both operations, as to be expected when drawing on a common memory network for multiplication and division. However, contrary to the revised identical elements model, De Brauwer and Fias (2009) found that although the children's skill level increased, the strongly interconnected developmental trajectories of both operations (as measured by a similar time course of problem size, five, and tie effects) did not decrease over time. This was surprising because fourth graders in the study of De Brauwer and Fias (2009) were quite skilled in solving multiplication and division problems with error rates around 6% and reaction times below 2 s.

These converging findings might be explained by the fact that children still get faster in solving division problems after primary school (Robinson et al., 2006). Thus, the children in the study of De Brauwer and Fias (2009) might not have reached the peak of their proficiency. As a consequence, the common network for multiplication and division might only diverge in later developmental stages, namely in older children in secondary school. Moreover, developmental trajectories might be different for children with different skill levels in solving multiplication and division problems. Furthermore, the relationship between simple

and complex multiplication and division problems has not yet been studied systematically.

In the present study, we aimed to investigate whether the relationship of multiplication and division after primary school depends on skill level. Skill level was indicated both by grade level (5th vs. 6th grade), school type (general or intermediate secondary school) as well as performance in arithmetic tasks. Grade level was used to investigate whether age-related differences in skill level modulate the relationship between multiplication and division. Therefore, we compared the relationship between multiplication and division in fifth grade to the relationship in sixth grade. The German educational system allowed us to divide children into groups of different skill levels depending on their secondary school type. This was possible because German children attend either one of three secondary school types based on their scholastic achievements in primary school: (1) general secondary school ("Hauptschule"), (2) intermediate secondary school ("Realschule"), or (3) grammar school ("Gymnasium"). The lowest achieving children usually attend general secondary school, the average achieving children attend intermediate secondary school, and the highest achieving children attend grammar school. Therefore, children's skill levels in multiplication and division problems should be higher in intermediate than in general secondary schools.

Overall, we expected strong developmental parallels between multiplication and division performance in line with the findings of De Brauwer and Fias (2009), that should be reflected in significant positive correlations between the two arithmetic operations. We expected a high correlation between simple multiplication and simple division, because of the usage of the division by mediation strategy by fifth and sixth graders (Robinson et al., 2006). Nevertheless, simple and complex operations should be interrelated because simple multiplications as well as divisions are sub-steps in deriving the result of the complex problems. This should be reflected by positive moderate to strong correlations. Moreover, the correlation between simple and complex division should be present even after controlling for performance in simple multiplication, because knowledge about simple division should be also relevant for solving complex division problems. Furthermore, according to the revised identical elements model (Rickard, 2005) we hypothesize lower correlations between simple multiplication and simple division performance for children with relatively higher skill levels (i.e., intermediate secondary school children, sixth graders and high performers) than children with lower skill levels (i.e., general secondary school children, fifth graders and low performers). We expect such a relationship, because for participants with low or intermediate skill levels the revised identical elements model assumes that participants rely on the division by mediation strategy. Therefore, we would expect that for participants with low or intermediate skill levels performance in simple multiplication and division should be correlated significantly. However, for participants with high skill levels the revised identical elements model assumes independent representations of multiplication and division facts. Therefore, performance of highly skilled participants in multiplication should be related less directly to performance in division resulting in relatively lower correlations than the correlation for

participants with low or intermediate skill levels who are applying the division by mediation strategy.

Because of this relationship and assuming that solving complex division problems builds on solving simple division problems, we furthermore expect relatively lower correlations between simple multiplication and complex division performance for children with relatively higher skill levels than with lower skill levels. As a consequence, this pattern should also be present for correlations between simple and complex division problems. Finally, we were also interested as to whether the relationship between simple and complex multiplication depends on skill level. However, the current state of research does not allow for a directed hypothesis on this issue.

Moreover, these correlations might be influenced by other cognitive abilities. Because some children are cognitively more advanced than others, performances across many cognitive tasks are usually correlated in children. To identify the unique contribution of performance in multiplication tasks to performance in division tasks, we controlled for the influence of non-verbal intelligence and short-term memory on these interrelations. Moreover, significant correlations between multiplication and division might only suggest that arithmetic abilities are generally interrelated. Therefore, we also used children's performance in addition and subtraction tasks as a control variable to identify the specific partial (co-) variance, shared only by multiplication and division.

MATERIALS AND METHODS
PARTICIPANTS

We assessed the performance of 392 students in fifth and sixth grade, in two different school types of the German educational system (general secondary school, i.e., "Hauptschule," and intermediate secondary school, i.e., "Realschule"). Students were recruited from 20 classes of nine different schools located in urban and suburban areas surrounding the city of Tuebingen (Germany) with mainly middle-class neighborhoods. We assessed the same number of classes from each school type (ten classes each). Because class sizes were smaller in general secondary schools, this resulted in a smaller sample for general than for intermediate secondary schools. In total our sample comprised 76 fifth graders (35 female, 41 male; mean age = 11.36 years, $SD = 0.49$ years) and 75 sixth graders (31 female, 44 male; mean age = 12.50 years, $SD = 0.63$ years) attending general secondary school and 112 fifth graders (56 female, 56 male; mean age = 10.88 years, $SD = 0.41$ years) and 129 sixth graders attending intermediate secondary school (100 female, 29 male; mean age = 11.89 years, $SD = 0.45$ years). Parental consent was obtained prior to the study.

TASKS AND PROCEDURE

As part of a larger project[1], we assessed students' performance on several numerical and arithmetic tasks. Students completed simple and complex multiplication and division problems and easy

---

[1]In addition to the pen and paper tasks on arithmetic, we also collected data about students' proficiency in orthography. Furthermore, students also completed computerized numerical tasks which addressed research questions different than the ones presented here.

and difficult addition and subtraction problems. We also assessed children's non-verbal intelligence and verbal short-term memory.

Simple multiplication problems involved all possible single digit multiplications excluding tie problems (e.g., 3×3) and multiplications including "1" or "0" as one of the operands. These exclusions were made based on research indicating that ties and problems including "1" or "0" were easier or solved with different strategies (see Campbell and Graham, 1985; Cooney et al., 1988; Siegler, 1988; Campbell and Gunter, 2002). This resulted in a set of 56 simple multiplication problems (e.g., 3×4). The inverse problems of these multiplications were used to create the *simple division* problems, resulting in a comparable set of 56 division problems (e.g., $3 \times 4 = 12 \rightarrow 12 \div 4 = 3$).

In all *complex multiplication* problems, one of the factors was a two-digit number whereas the other one was either a single-digit or a two-digit number. Single-digit numbers ranged from 2 to 9 and two-digit numbers ranged from 12 to 19. Again, tie problems (e.g., "12×12") were excluded. Results were either two- or three-digit numbers. Of the total 28 complex multiplication problems, 10 consisted of a single-digit and a two-digit factor (e.g., 7×16), 9 of a two-digit and a single-digit factor (e.g., 15×6) and 9 of 2 two-digit factors (e.g., 14×17).

In *complex division* problems, dividends were either two-digit or three-digit numbers. The same restrictions as for complex multiplication problems were applied in creating the division problems. Divisors and quotients were always different numbers (inversely to tie problems in multiplication). In a complex division problem, either the divisor or the quotient was a single-digit number ranging from 2 to 9, whereas the other was a two-digit number ranging from 12 to 19. Of the total 28 complex division problems, 9 consisted of a two-digit dividend and a single-digit divisor (e.g., $95 \div 5$), 9 of a three-digit dividend and a single-digit divisor (e.g., $108 \div 6$), 8 of a two-digit dividend and divisor (e.g., $70 \div 14$), and 2 of a three-digit dividend and a two-digit divisor (e.g., $153 \div 17$).

We created 36 *easy* and 36 *difficult addition* problems. Whereas none of the easy addition problems involved a carry operation (e.g., $42 + 25$), half of the difficult addition problems required a carry operation (e.g., $69 + 18$). *Easy* and *difficult subtraction* problems comprised 30 problems each. The easy subtraction problems did not require a borrowing operation (e.g., $54 - 31$), whereas half of the difficult subtraction problems did (e.g., 63–17). Terms were single-digit, two-digit or three-digit numbers. Problems were ordered by increasing number of digits, starting with problems consisting only of single-digit numbers first and ending with problems consisting only of three-digit numbers.

Arithmetic tasks were administered to entire classrooms, as speeded paper and pencil tests with a strict time limit to prevent ceiling effects. Multiple problems were presented on each page as production tasks (e.g., "$2 \times 7 =$ __"; "$21: 3 =$ __") with students being instructed to solve as many problems as possible within the allotted time. Time limits were 1.5 min for simple multiplication and division tasks, 1.5 min for easy and difficult addition and subtraction tasks and 2 min for complex multiplication and division tasks. Students were not allowed to take notes. For each student, we calculated performance in simple and complex multiplication and in simple and complex division tasks as the number

of correctly solved problems. Performances in both easy and difficult addition and subtraction problems were summed up to get one composite performance score for ability in solving addition and subtraction problems.

Non-verbal intelligence was assessed by conducting the subtest "matrices" of the Culture Fair Test (CFT-20-R, Weiss, 2006). The CFT-20-R is supposed to assess fluid intelligence, which is the capacity to think logically and solve problems in novel situations, independent of acquired knowledge. Children were instructed to solve as many of the 17 items as possible within the allotted time of 3 min.

Verbal short-term memory was assessed with a verbal learning and memorizing test (i.e., "Verbaler Lern- und Merkfähigkeitstest," Helmstaedter et al., 2001). Fifteen words were consecutively read aloud to the classroom. After all words were presented, children were instructed to recall and write down as many words as possible within a time frame of 2 min.

ANALYSIS

Firstly, we evaluated whether performance indeed differed in the expected directions for students of different school types and of different grades. We assessed performance of children in simple and complex multiplication and division tasks. The MANOVA is the appropriate method to examine group differences when there are two or more dependent variables. Therefore, performance scores in simple and complex multiplication and division tasks were submitted to a 2×2 MANOVA with the independent variables school type (general secondary school vs. intermediate secondary school) and grade level (fifth vs. sixth grade) and the dependent variables performance in arithmetic tasks (simple and complex multiplication and simple and complex division).

Secondly, we computed bivariate correlations and partial correlations between performance in simple and complex multiplication and division tasks for the whole sample as well as separately for each grade level, school type and performance group. Performance groups were created by dividing our sample in low and high performance according to their skill in multiplication and division tasks using a median split[2]. Skill in multiplication and division tasks was calculated by adding children's performance scores in simple and complex multiplication and division tasks. Please note that we are fully aware of the problems of dichotomizing a continuous variable (e.g., Maccallum et al., 2002; Irwin and Mcclelland, 2003). However, dichotomizing a continuous variable can be interpreted as a more conservative test in case of bivariate analyses. Thus, if correlations are still present for low and high performers, this will indicate that the relationship between two variables is strong. Nevertheless, the median split analysis is only a supplementary analysis and findings should be interpreted with caution. In partial correlations, effects of non-verbal intelligence, verbal short-term memory and the addition and subtraction performance scores were removed. Resulting correlation coefficients were then compared applying Fisher's *r*-to-*z*-transformation in which Pearson's *r* values are converted into normally distributed *z* values. The Fisher *r*-to-*z*-transformation allows assessing the significance of the difference

---

[2] We wish to thank Reviewer Elena Ise for suggesting this analysis.

between two correlation coefficients found in two independent samples.

RESULTS

ANALYSES OF VARIANCE

In general, students' performance varied depending on grade level and school type. Both main effects of grade level and school type were significant [school type: Pillai-trace = 0.13, $F_{(4, 385)} = 13.71$, $p < 0.001$, $\eta_p^2 = 0.13$; grade: Pillai-trace = 0.13, $F_{(4, 385)} = 14.12$, $p < 0.001$, $\eta_p^2 = 0.13$]. The interaction between school type and grade was not significant [Pillai-trace = 0.01, $F_{(4, 385)} = 1.40$, $p = 0.23$, $\eta_p^2 = 0.01$]. Subsequent univariate analyses indicated that main effects of grade level and school type were significant in both arithmetic tasks on both difficulty levels (see **Table 1**; for means and SDs of control variables see **Table A1**; for min/max scores see **Table A2**). In all four tasks, sixth graders outperformed fifth graders and students from intermediate secondary schools outperformed students from general secondary schools. Thus, students' performance differed in the expected direction.

CORRELATIONS

When calculating correlations between all tasks for the whole sample, we found moderate to strong bivariate and partial correlations between all tasks, indicating that performance in all tasks was related (see **Table 2**). Moreover, after adding performance in simple multiplication as a covariate, the partial correlation between simple and complex division was still significant. It did, however, decrease from $r_{(387)} = 0.38$ to $r_{(386)} = 0.27$.

We then tested whether the correlation between multiplication and division performance depended on children's skill level as reflected by grade level and school type. Accordingly, we calculated partial correlations between tasks separately for each grade level, school type, and performance group and compared them using Fisher's r-to-z-transformation (see **Table 3**).

We observed reliably higher partial correlations between simple multiplication and simple division and between simple multiplication and complex multiplication for students from intermediate secondary schools than for students from general secondary schools. A similar pattern was present for performance groups. Partial correlations between simple multiplication and simple division were higher for higher-skilled students than lower-skilled students. However, we found no differences in partial correlation coefficients between simple multiplication and simple division for age groups.

Results for the partial correlations between simple and complex operations were different for grade level, school type and performance in multiplication and division tasks. Partial correlations between simple multiplication and complex division and between simple division and complex division differed between students from general and intermediate secondary school. As hypothesized, the partial correlation was higher for students from intermediate secondary school than from general secondary school. For performance groups, we found a significant difference for partial correlations between simple and complex multiplications with a higher correlation for high-skilled students than for low-skilled students. Other partial correlation coefficients did not differ. Moreover, we did not find any significant differences between age groups.

DISCUSSION

In the present study, we aimed to explore whether the strong developmental parallels between multiplication and division performance persist into secondary school and to what extent this depends on skill level. Therefore, we assessed the performance of fifth and sixth graders of two secondary school types (general vs. intermediate secondary school) of the German educational system in multiplication and division problems. In line with our hypotheses, we found that sixth graders outperformed fifth graders and students from intermediate secondary schools outperformed students from general secondary schools. Thus, skill levels differed both between age groups (fifth vs. sixth graders)

Table 2 | Bivariate and partial correlations (controlling for non-verbal intelligence and verbal short-term memory) between arithmetic tasks.

	Complex multiplication	Simple division	Complex division
BIVARIATE CORRELATIONS			
Simple multiplication	0.59**	0.78**	0.45**
Complex multiplication		0.62**	0.56**
Simple division			0.53**
PARTIAL CORRELATIONS			
Simple multiplication	0.43**	0.70**	0.29**
Complex multiplication		0.45**	0.41**
Simple division			0.38**

Bivariate correlations: df = 390, partial correlations: df = 388, **p < 0.001.

Table 1 | Means (SD in parenthesis) and F-values for general (GS) and intermediate (IS) secondary schools and fifth and sixth grade.

Task	Grade				School type			
	5.	6.	F	η_p^2	GS	IS	F	η_p^2
Simple multiplication	14.79 (7.07)	20.45 (9.03)	46.15**	0.10	15.23 (8.13)	19.30 (8.56)	22.01**	0.05
Complex multiplication	5.44 (3.17)	7.23 (3.55)	25.91**	0.06	5.13 (3.26)	7.15 (3.41)	33.54**	0.08
Simple division	16.65 (8.21)	22.57 (10.37)	39.13**	0.09	15.79 (8.45)	22.20 (9.86)	44.67**	0.10
Complex division	2.43 (2.28)	4.05 (3.05)	28.66**	0.07	2.36 (2.31)	3.84 (2.97)	27.24**	0.07

df1 = 1, df2 = 388; **p < 0.001.

Table 3 | Partial correlations (controlling for non-verbal intelligence, verbal short-term memory and performance in solving addition and subtraction tasks) between arithmetic tasks separately for fifth and sixth grades, general (GS) and intermediate (IS) secondary schools and low-skilled and high-skilled students.

Task 1	Task 2	Grade			School type			Performance group		
		5.	6.	z	GS	IS	z	Low	High	z
Simple multiplication	Simple division	0.62	0.71	−1.59	0.62	0.74	−2.04*	0.32	0.64	−4.12**
Simple multiplication	Complex division	0.16	0.27	−1.15	0.12	0.35	−2.37**	−0.08	0.06	−1.34
Simple multiplication	Complex multiplication	0.36	0.42	−0.69	0.39	0.44	−0.62	0.02	0.25	−2.33**
Simple division	Complex division	0.30	0.36	−0.71	0.20	0.44	−2.49**	0.14	0.16	−0.24

Partial correlations were compared using Fisher's r-to-z transformation.

GS: df = 146; IS: df = 236; 5. grade: df = 183; 6. Grade: df = 199; low-skilled: df = 191; high-skilled: df = 191; $^*p < 0.05$, $^{**}p < 0.01$(one-sided).

as well as between school types (general vs. intermediate secondary school). In the following section, we will first discuss age-related differences before elaborating on school type differences and finally, we will discuss the specificity of correlations between multiplication and division tasks.

AGE-RELATED INFLUENCES

Our results indicated that strong developmental parallels between multiplication and division persist into fifth and sixth grade of secondary school. In particular, simple multiplication and division were strongly interrelated, even after controlling for non-verbal intelligence, verbal short-term memory and arithmetic performance in addition and subtraction tasks. Thus, we suggest that, comparable to the study of Robinson et al. (2006) students mostly relied on the division by mediation strategy when solving division problems. According to Robinson et al. (2006), fifth and sixth graders applied this strategy in about 50% of all problems, making it the most applied strategy in these age groups. Extending the results of De Brauwer and Fias (2009), who found that simple multiplication and division were interrelated until grade 4, we found evidence that they are still interrelated reliably in grade 6.

Furthermore, simple and complex tasks were interrelated reliably, suggesting that knowledge about simple problems is indeed recruited for solving complex problems. More specifically, we found that students' knowledge about simple division contributed to their performance in solving complex division problems even after controlling for performance in simple multiplication.

However, bivariate and partial correlations were lower for complex tasks, which might be caused by children's overall lower performance in complex tasks. Thus, lower correlations might be due to floor effects reducing variability. Nevertheless, performance in complex tasks varied within a sufficiently large range (i.e., 0–13 correctly solved problems). This speaks against a general floor effect for complex problems. However, it would be interesting to evaluate the current results by giving children more time for the complex tasks allowing them to solve more problems.

Comparable to the study of De Brauwer and Fias (2009), our findings cannot be accounted for by the revised identical elements model (Rickard, 2005). According to the model, only students at low and intermediate skill levels should apply the division by mediation strategy. Contrary to this hypothesis, all students

were quite skilled at solving multiplication and division problems as indicated by their fast solution times. For multiplication problems, fifth graders took about 6.26 s per problem and sixth graders took about 4.50 s per problem. For division problems, fifth graders took about 5.35 s per problem and sixth graders took about 4.13 s per problem[3] . They were somewhat slower than students in the studies of Robinson et al. (2006) and De Brauwer and Fias (2009), who needed less than 3 s to solve multiplication and division problems. However, in the study by Robinson et al. (2006), students had to answer verbally and in the study of De Brauwer and Fias (2009), verification and number-matching tasks were used. In our study, students had to write down answers, which obviously took more time than responding verbally or pressing a button. Interestingly, students in our study were faster at solving simple division than simple multiplication problems, which seems to contradict our implication that they relied on an indirect strategy. However, again this finding can be explained by the nature of the tasks we used. Most simple multiplication problems required students to write down two digits, whereas all simple division problems required them to write down only one digit. Thus, slightly faster solution times for simple division problems might be due to less time needed to write down the solutions in simple division compared to simple multiplication problems.

Importantly, we did not find significant differences in correlation coefficients between fifth and sixth graders. Thus, we did not find age-related differences in skill level to modulate the relationship between multiplication and division. We assume that high correlations between simple multiplication and division tasks indicated that students relied on the division by mediation strategy. Thus, the reliance on the division by mediation strategy cannot explain age-related performance differences, because the observed correlations seem to indicate that in both age groups children relied on the strategy to a similar extent.

SCHOOL TYPE DIFFERENCES

Performance differences between students from general and intermediate secondary schools allowed us to look at how the relationship between multiplication and division depends on differences in skill level between secondary school types. We found that

---

[3]Time spent per problem was calculated by dividing the number of solved problems by the total allotted time.

students from intermediate secondary schools outperformed students from general secondary schools. In line with the revised identical elements theory (Rickard, 2005), we hypothesized that correlations between simple multiplication and division tasks should be lower for intermediate secondary school children than for general secondary school children. However, we found the exact opposite. Partial correlations were even higher for students from intermediate secondary schools, suggesting that their multiplication and division memory network was even more closely related. This interpretation was corroborated by similar findings for the partial correlation between simple multiplication and complex division. Solving complex division problems involves splitting them up into simple division problems. If students who are better at solving multiplication problems apply the mediation by division strategy, they should also be better at solving complex division problems. Thus, our results suggest that students from intermediate secondary schools who performed better in simple and complex division relied more frequently or used conceptually or procedurally more consistently the inverse relationship between multiplication and division.

The relationship between skill level and simple and complex problems was less clear. For the correlation between simple multiplication and complex multiplication we found that skill level as indexed by differences between school types had no influence. Nevertheless, the significant correlations between simple and complex multiplication indicate that students from general as well as from intermediate secondary schools rely on their knowledge about simple multiplications when solving complex multiplications. Thus, our finding suggests that students from different school types rely on their knowledge about simple multiplication to a similar amount. However, students from intermediate schools did not seem to rely more heavily on their knowledge about simple multiplications.

In contrast to multiplication, we found that skill level as indexed by differences between school types moderated the correlations between simple and complex division. Thus, it seems that students with higher skill levels (i.e., students from intermediate secondary schools) relied more heavily on their skills in simple division to solve complex division problems. However, we found a similar pattern between simple multiplication and simple division and between simple multiplication and complex division. Therefore, one possible interpretation is that students from intermediate secondary schools made greater use of the division by mediation strategy in both simple and complex division problems by retrieval of simple multiplication facts.

PERFORMANCE DEPENDENT DIFFERENCES

We also examined whether correlations between multiplication and division tasks depended on students' performance in multiplication and division tasks. Therefore, we allocated students to one of two performance groups.

In accordance with findings for different school types, we found higher correlations for more skilled students. Higher correlations suggest that more skilled students seemed to rely more heavily on the inverse relationship between simple multiplication and division, and thus, they used the division by median strategy more often. Thus, we found confirmatory evidence that skill

level modulates the relationship between simple multiplication and division in secondary school children.

For the relationship between simple and complex tasks, findings for performance groups differed from findings for school types. Correlations between simple multiplication and complex division and between simple division and complex division were not modulated by skill level. Thus, more skilled students did not seem to rely more heavily on their knowledge about simple multiplication and simple division when solving complex division problems. Moreover, correlation coefficients between simple multiplication and complex division were close to zero suggesting that students do not rely on their knowledge about simple multiplication when solving complex division problems. However, there was a reliable partial correlation between simple multiplication and complex division before allocating students to different groups [$r_{(386)} = 0.29, p < 0.001$]. Thus, the disappearance of this association may be explained by the method used to create two performance groups. It has been shown repeatedly that a median split can reduce correlation coefficients between two variables as it dichotomizes a continuous variable and thus reduces variability which is detrimental in correlation analyses (Maccallum et al., 2002; Irwin and Mcclelland, 2003). In the present case, this might have led to the reduced correlation between simple multiplication and complex division.

Furthermore, we observed that skill level modulated the relationship between simple and complex multiplication. Correlation coefficients were higher for more skilled students suggesting that they relied more on their knowledge about simple multiplications when solving complex multiplications. Moreover, a near to zero correlation coefficient for the low performance group seems to imply that students from this group did not rely on their knowledge about simple multiplication when solving complex multiplication problems. Because knowledge about simple multiplication is necessary for solving complex multiplications, this finding is quite implausible and might again stem from creating dichotomous performance groups from a continuous variable by means of a median split.

Taken together, the performance group analysis confirmed our finding of skill related differences as indexed by differences between school types for the relationship between simple multiplication and division. Moreover, contrary to the revised identical elements model (Rickard, 2005) we did not find any evidence that the relationship between simple and complex multiplication and division tasks was weaker for high performers than low performers.

SPECIFICITY OF CORRELATIONS

In this study, we used general cognitive and arithmetic abilities as control variables to identify the unique covariance between performance in multiplication and division tasks. All partial correlations remained significant after controlling for intelligence, short-term memory and general arithmetic ability (as measured by performance in addition and subtraction tasks). However, they were lower than the bivariate correlations suggesting that general cognitive and arithmetic capabilities contribute to bivariate correlations between multiplication and division. Thus, bivariate correlations between these operations may overestimate the

unique covariance of both operations, because better cognitive and arithmetic capabilities may lead to better performances in both operations.

The partial correlations allowed us to identify which tasks are most closely interrelated. After controlling for general cognitive and arithmetic abilities, the correlation between simple multiplication and simple division was higher than all other correlations. Complying with the findings of Robinson et al. (2006), one possible interpretation of this finding is that fifth and sixth graders mostly rely on their knowledge about multiplication facts to solve simple division problems. In contrast, we found lower correlations between simple multiplication and complex multiplication and division tasks after controlling for general cognitive and arithmetic abilities. Hence, general cognitive and arithmetic abilities as well as procedural knowledge might be more important than knowledge about multiplication facts when solving complex tasks.

However, a possible limitation of our study is that we only inferred strategies used from correlations we observed between multiplication and division performance. While our results provide conclusive evidence that children seemed to make use of the division by mediation strategy, a more direct investigation of their strategy use would be beneficial. One way to do so might be the consideration of verbal reports as used by Robinson et al. (2006). Moreover, correlations do not allow for causal conclusions. Therefore, even if it seems implausible from a developmental point of view (division is introduced after multiplication), children might have relied on their knowledge about simple division when solving simple multiplication problems.

CONCLUSIONS

In the current study we were interested (i) whether the close developmental parallels between multiplication and division persist beyond primary school, (ii) whether similar developmental parallels can be observed for simple and complex problems, (iii) whether these relationships are influenced by skill level, and (iv)

whether the correlations are specific and not driven by general cognitive or arithmetic abilities. By collecting data of fifth and sixth graders attending general or intermediate secondary schools we operationalized skill level both within as well as between age groups. In general, our findings provide converging evidence for the importance of the inverse relationship between multiplication and division. In line with this interpretation, moderate to strong partial correlations between multiplication and division performance - even after controlling for short-term memory, intelligence and other arithmetic abilities, indicate that students seem to recast division problems as multiplication problems. Importantly, skill level as indexed by school type influenced this relationship, as students from intermediate secondary schools seem to draw more heavily on multiplication fact retrieval in division as indicated by higher correlations between the two operations. This finding was confirmed for the relationship between simple multiplication and division when analyzing performance dependent difference. Yet, we did not find any evidence that age-related differences in skill level modulated the relationship between multiplication and division. These findings are hard to reconcile with the predictions made by the identical elements model, which claims that mediation of division by multiplication should decrease as skill level increases.

REFERENCES

Butterworth, B., Zorzi, M., Girelli, L., and Jonckheere, A. R. (2001). Storage and retrieval of addition facts: the role of number comparison. *Q. J. Exp. Psychol.* 54, 1005–1029. doi: 10.1080/02724980143000064

Campbell, J. I., and Gunter, R. (2002). Calculation, culture, and the repeated operand effect. *Cognition* 86, 71–96. doi: 10.1016/S0010-0277(02)00138-5

Campbell, J. I. D., and Graham, D. J. (1985). Mental multiplication skill - structure, process, and acquisition. *Can. J. Psychol.* 39, 338–366. doi: 10.1037/h0080065

Cooney, J. B., Swanson, H. L., and Ladd, S. F. (1988). Acquisition of mental multiplication skill - evidence for the transition between counting and retrieval strategies.

Cogn. Instr. 5, 323–345. doi: 10.1207/s1532690xci0504_5

De Brauwer, J., and Fias, W. (2009). A longitudinal study of children's performance on simple multiplication and division problems. *Dev. Psychol.* 45, 1480–1496. doi: 10.1037/a0015465

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239

Geary, D. C., and Widaman, K. F. (1987). Individual-differences in cognitive arithmetic. *J. Exp. Psychol. Gen.* 116, 154–171. doi: 10.1037/0096-3445.116.2.154

Heath, T. L. (1956). *The Thirteen books of the Elements of Euclid,* Vol. 2 (Books III–IX). New York, NY: Dover.

Helmstaedter, C., Lendt, M., and Lux, S. (2001). *Verbaler Lern- und Merkfähigkeitstest.* Göttingen: Hogrefe Verlag.

Hickendorff, M., Van Putten, C. M., Verhelst, N. D., and Heiser, W. J. (2010). Individual differences in strategy use on division problems: mental versus written computation. *J. Educ. Psychol.* 102, 438–452. doi: 10.1037/a0018177

Hope, J. A., and Sherrill, J. M. (1987). Characteristics of unskilled and skilled mental calculators. *J. Res. Math. Educ.* 18, 98–111. doi: 10.2307/749245

Irwin, J. R., and Mcclelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *J. Mark. Res.* 40, 366–371. doi: 10.1509/jmkr.40.3.366.19237

Lemaire, P., and Siegler, R. S. (1995). Four aspects of strategic change:

contributions to children's learning of multiplication. *J. Exp. Psychol. Gen.* 124, 83–97. doi: 10.1037/0096-3445.124.1.83

Maccallum, R. C., Zhang, S., Preacher, K. J., and Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychol. Methods* 7, 19–40. doi: 10.1037/1082-989X.7.1.19

Raghubar, K. P., Barnes, M. A., and Hecht, S. A. (2010). Working memory and mathematics: a review of developmental, individual difference, and cognitive approaches. *Learn. Individ. Dif.* 20, 110–122. doi: 10.1016/j.lindif.2009.10.005

Rickard, T. C. (2005). A revised identical elements model of arithmetic fact representation. *J. Exp. Psychol. Learn. Mem. Cogn.* 31, 250–257. doi: 10.1037/0278-7393.31.2.250

Rickard, T. C., and Bourne, L. E. (1996). Some tests of an identical elements model of basic arithmetic skills. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 1281–1295. doi: 10.1037/0278-7393.22.5.1281

Rickard, T. C., Healy, A. F., and Bourne, L. E. (1994). On the cognitive structure of basic arithmetic skills - operation, order, and symbol transfer effects. *J. Exp. Psychol. Learn. Mem. Cogn.* 20, 1139–1153. doi: 10.1037/0278-7393.20.5.1139

Robinson, K. M., Arbuthnott, K. D., Rose, D., McCarron, M. C., Globa, C. A., and Phonexay, S. D. (2006). Stability and change in children's division strategies. *J. Exp.*

Child Psychol. 93, 224–238. doi: 10.1016/j.jecp.2005.09.002

Seitz, K., and Schumann-Hengsteler, R. (2000). Mental multiplication and working memory. *Eur. J. Cogn. Psychol.* 12, 552–570. doi: 10.1080/095414400750050231

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *J. Exp. Psychol. Gen.* 117, 258–275. doi: 10.1037/0096-3445.117.3.258

Verguts, T., and Fias, W. (2005). Interacting neighbors: a connectionist model of retrieval in single-digit multiplication. *Mem. Cogn.* 33, 1–16. doi: 10.3758/BF03195293

Weiss, R. (2006). *Grundintelligenztest (CFT 20-R)*. Göttingen: Hogrefe.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology.

APPENDIX

Table A1 | Means (SD in parenthesis) of control variables for general (GS) and intermediate (IS) secondary schools and fifth and sixth grade.

Task	School type		Grade	
	GS	*IS*	5.	6.
Verbal short-term memory	6.99 (2.35)	8.07 (1.77)	7.40 (1.95)	7.89 (2.17)
Non-verbal intelligence	8.26 (2.63)	8.92 (2.73)	8.60 (2.60)	8.72 (2.81)
Score in addition and subtraction	45.79 (9.94)	51.29 (8.73)	47.64 (9.22)	50.58 (9.72)

Table A2 | Minimum (min), maximum (max), mean and SD of each variable.

	Min	Max	Mean	*SD*
Simple multiplication	1	51	17.73	8.62
Complex multiplication	0	17	6.37	3.49
Simple division	0	55	19.73	9.84
Complex division	0	13	3.27	2.82
Verbal short-term memory	0	14	7.65	2.08
Non-verbal intelligence	0	14	8.67	2.71
Score in addition and subtraction	16	75	49.17	9.58

Neural correlates of the numerical distance effect in children

Christophe Mussolin[1]*, Marie-Pascale Noël[2], Mauro Pesenti[2,3], Cécile Grandin[3] and Anne G. De Volder[3,4]

[1] Laboratory Cognition, Language, Development, Center for Research in Cognition and Neurosciences, Université Libre de Bruxelles, Brussels, Belgium
[2] Centre de Neurosciences Système et Cognition, Institut de Recherche en Sciences Psychologiques, Université Catholique de Louvain, Louvain-la-Neuve, Belgium
[3] School of Medicine, Institute of Neuroscience, Université Catholique de Louvain, Brussels, Belgium
[4] Pediatric Neurology Service, Cliniques Universitaires St. Luc, Brussels, Belgium

In number comparison tasks, the performance is better when the distance between the numbers to compare increases. It has been shown that this so-called numerical distance effect (NDE) decreases with age but the neuroanatomical correlates of these age-related changes are poorly known. Using functional magnetic resonance imaging (fMRI), we recorded the brain activity changes in children aged from 8 to 14 years while they performed a number comparison task on pairs of Arabic digits and a control color comparison task on non-numerical symbols. On the one hand, we observed developmental changes in the recruitment of frontal regions and the left intraparietal sulcus (IPS), with lower activation as age increased. On the other hand, we found that a behavioral index of selective sensitivity to the NDE was positively correlated with higher brain activity in a right lateralized occipito-temporo-parietal network including the IPS. This leads us to propose that the left IPS would be engaged in the refinement of cognitive processes involved in number comparison during development, while the right IPS would underlie the semantic representation of numbers and its activation would be mainly affected by the numerical proximity between them.

Keywords: number magnitude, children, intraparietal sulcus, frontal cortex, development

INTRODUCTION

Being able to compare two numbers according to their magnitudes is a prerequisite to acquire and master mathematical abilities during childhood. In this kind of task, latencies and error rates decrease as the numerical difference between two numbers to be compared increases (Moyer and Landauer, 1967). The classical interpretation of this so-called numerical distance effect (NDE) is that numbers are automatically converted into an internal representation like analog magnitudes that are in turn compared with each other [but see Verguts and Fias (2004), for a different view]. Since this seminal observation, the NDE has been replicated and manipulated in a large body of adult studies.

Only a few behavioral studies investigate how the NDE evolves during development and the published findings are unclear. Sekuler and Mierkiewicz (1977) assessed performance of children from kindergarten, first-, fourth-, and seventh-grade as well as performance of adult students while they had to compare pairs of Arabic digits. Although the NDE affected performance in all age groups, as reflected by shorter latencies for far distances relative to close distances, the difference in processing speed was more pronounced for younger participants than for older ones. This finding was also reported in other experiments (Duncan and McFarland, 1980; Holloway and Ansari, 2009). The age-related change in the size of NDE could be interpreted as a greater efficiency to access or represent number magnitude as the age increases. Alternatively, it is also possible that the flatter NDE with age corresponds to gradually faster domain-general speed of

processing. Indeed, similar developmental changes in the distance effect were observed in number, brightness, and height comparisons (Holloway and Ansari, 2008), suggesting that at least some of the cognitive mechanisms involved in the NDE may be shared by other quantity dimensions.

With the emergence of functional neuroimaging techniques, there were numerous attempts to define which brain areas are responsible for the NDE in adults. Using positron emission tomography (PET) or functional magnetic resonance imaging (fMRI) these experiments provided demonstration that brain areas in and around the intraparietal sulcus (IPS) in both hemispheres were consistently activated during number comparison (Pinel et al., 1999; Pesenti et al., 2000), even in the absence of explicit processing of number magnitude (Eger et al., 2003). Crucially, the extent of brain activation in these cortical areas was found to be modulated by the NDE. Indeed, the brain activation level in the left and right IPS was higher when numerical distance was smaller, both for Arabic numbers and number words (Pinel et al., 2001; Notebaert et al., 2010) as well as for non-symbolic numerosities (i.e., collections of dots, Piazza et al., 2004; Ansari et al., 2006; Cantlon et al., 2006). These findings indicate that areas around the IPS might be the neural substrate of number magnitude representation in adults (Dehaene et al., 2003).

Several neuroimaging studies aimed to examine the developmental course in brain activity changes related to basic numerical abilities such as Arabic number comparison or numerosity habituation. However, current data are mixed. On the one hand,

several right-lateralized parietal regions including the IPS were more strongly modulated by NDE in adults than in 10-year-old children during symbolic number comparison. Inversely, the modulation effect in the right precentral and inferior frontal gyri was stronger in children than in adults (Ansari et al., 2005). It is noteworthy that these results were only partially replicated with a non-symbolic comparison task (Ansari and Dhital, 2006). On the other hand, when adults and 4-year-old children were habituated to numerosities varying either on number or element shape, the IPS activation changes linked to number changes overlapped considerably in both groups (Cantlon et al., 2006). This is in accordance with previous observations using evoked related potentials (ERPs) showing that NDE was associated with similar components, localization and timing in 5-year-old children and adults who compared canonical dot patterns or Arabic digits (Temple and Posner, 1998).

Altogether, behavioral data consistently pointed out a weaker sensitivity to NDE for older participants compared to younger— i.e., reflected by a smaller difference between close and far numerical distances. Several neuroimaging studies also pointed out developmental changes related to NDE reflected by a progressive recruitment of parietal regions with age, coupled with a decreasing engagement of frontal regions. However, it is not clear whether these age-related changes correspond to a refinement of the number magnitude representation or rather an improvement in other processes involved in the NDE such as the access to the magnitude from symbolic numbers. To distinguish between these proposals, it is necessary to examine, amongst brain regions modulated by the NDE, which of them are specifically dedicated to number magnitude processing and whether or not the activity in these regions is also modulated by age. Indeed, as pointed out by different fMRI studies in the past (Pinel et al., 2001, 2004; Cohen Kadosh et al., 2005), not only the regions in and around the IPS showed a modulation of their activity by the NDE but also other regions that are not supposed to play a role in the number magnitude representation. To date, no clear explanation has been provided concerning the role played by these regions outside the parietal cortex in the context of number processing.

The present fMRI study aims at addressing these questions by recording neural responses in children aged from 8 to 14 years who were required to compare two Arabic digits separated by either a close or a far distance. This age range was chosen to ensure that children had sufficient knowledge with Arabic digits to compare the magnitude they convey with a high accuracy (Sekuler and Mierkiewicz, 1977; Holloway and Ansari, 2009). In the current study, we paid special attention to include a color comparison task on pairs of non-numerical symbols whose colors were either close (e.g., red vs. pink) or far (e.g., red vs. blue) from each other[1]. It is important to include such kind of control task in order to tap brain activation changes specifically related to number magnitude processing since several adult experiments have pointed out a recruitment of IPS in non-numerical quantity comparison

(Pinel et al., 2004; Cohen Kadosh et al., 2005). Accordingly, subtracting the activity related to color distances from the activity related to numerical distances should allow us to point out which brain regions are recruited by the different steps of number comparison (i.e., accessing numerical symbols, mapping these symbols to the number magnitude representation, and activating these magnitudes), while excluding the common cognitive processes engaged in both tasks (e.g., processing visual inputs, performing a comparison and selecting a response). In addition, to disentangle which brain areas, amongst the activated ones, are involved in the NDE (i.e., isolating those whose activation is dedicated to number magnitude processing from those whose activity is not), we also computed a behavioral index of sensitivity to numerical proximity (dRT score) for each child (see Materials and Methods). We used individual (dRT) scores in a correlation analysis to examine the effect of (im)precision of the number magnitude representation on brain activation patterns. By analyzing correlations between brain activity and age or dRT scores, we aimed to get further insight into the cognitive processes engaged in the NDE during development. A common correlation between IPS activity and age and dRT scores (in opposite directions[2]) would indicate that the age-related changes in the size of NDE observed in behavioral studies reflect a refining of the number magnitude representation during development. Alternatively, if the regions modulated by age differ from those modulated by dRT scores, this would lead us to postulate that the behavioral improvement associated with the changes in the size of NDE with age does not tap a refining of the number magnitude representation *per se*, but rather would refer to less specific processes playing a role in the access to this representation or in the comparison between magnitudes.

MATERIALS AND METHODS

PARTICIPANTS

Nineteen children aged from 8 to 14 years (6 girls, 2 left-handed, average age 10.5 ± 1.7 years) participated in the fMRI experiment. All participants had normal or corrected-to-normal vision. They were healthy and medication-free with no history of neurological illness or learning disability. All protocols were approved by the local ethics committee of the UCL school of medicine, and have been conducted according to the principles expressed in the Declaration of Helsinki. All legal guardians of the children gave informed written consent prior to the experiment.

EXPERIMENTAL TASKS

During the fMRI sessions, participants performed two tasks each comprising two distance levels, giving rise to four conditions presented in separate blocks. In the number comparison task, two Arabic digits from 1 to 9, separated by close (1 or 2) or far (5 or 6) numerical distances were presented to the participants. In the color comparison task, two non-numerical symbols (selected

---

[1]In our control task we have chosen to use non-numerical symbols and to avoid colored Arabic digits in order to prevent automatic activation of number magnitude related to them (Girelli et al., 2000; Rubinsten et al., 2002), which could in turn lead to brain activation around the IPS (Pinel et al., 2004).

[2]In this case, positive correlations between dRT and IPS activity are expected since a high sensitivity to numerical proximity should be associated with a larger activation in brain regions dedicated to number magnitude processing. By contrast, IPS regions should be negatively correlated with age since a lesser requirement of these regions is expected as experience with number magnitudes increases.

amongst the symbols Λ, Δ, Ω, Σ, Γ, δ, &, Φ) were presented. The target symbol was red and the color of the other symbol was either close to (i.e., pink) or far from (i.e., blue) red. The pairs of Arabic digits or non-numerical symbols appeared every 1800 ms. Stimuli were flashed for 200 ms at the center of the screen, on both side of a fixation point, followed by a fixed 1600 ms interval. The participants were instructed to keep their eyes on the fixation point throughout the experiment and to avoid movements as much as possible. The participants held a MRI-compatible response button in each hand, and had to select the larger digit (numerical task) or the red symbol (color task) of each pair by pressing the corresponding left or right button as quickly as possible. The position of the correct response was counterbalanced for each trial. Response latencies were measured from the disappearance of the stimuli. During the fixation periods, participants were asked to look at the fixation point without making head or eye movement. Stimuli were presented using a video projector and a translucent screen. The experiment was programmed and responses recorded using E-Prime 1.2 software (Schneider et al., 2002).

Before the scanning session, all participants were carefully instructed about the whole procedure and had to solve four blocks of practice trials outside the MRI room to familiarize with the tasks.

NEUROPSYCHOLOGICAL TASKS

Children's general cognitive abilities were assessed as follows. Intellectual capacities were evaluated using the Similarities and Images Completion subtests of the WISC-III (Wechsler, 1996) which enabled to calculate an estimate of IQ for each child. Four measures of short-term memory were also obtained. In the word span tasks, children were presented with increasingly longer series of words and were asked to repeat them in the actual presentation order (forward word span) or in the reverse order (backward word span). The Corsi block-tapping test (Corsi, 1972) provided a measure of spatial short-term memory. In this task, children were asked to reproduce the same sequence of block pointing as shown by the examiner. The listening span [adapted from Daneman and Carpenter (1980)] was used to evaluate the central executive component of the working memory. In this test, the experimenter read sets of sentences (from two to four) and the child was required to indicate whether each sentence was true or not. Then, at the end of the set, the child had to recall the last word of each of the sentences included in the set. Individual scores in these neuropsychological tasks were in the normal range. A summary of individual results is provided in Supplementary Table S1.

BEHAVIORAL DATA ANALYSIS

An individual measure of the sensitivity to numerical proximity was obtained following an approach adapted from that of Holloway and Ansari (2009). For each child, reaction times (RTs) for comparisons with far (median of distances 5 and 6) distances were subtracted from those with close (median of distances 1 and 2) numerical (or color) distances. These values were then divided by the RTs for far distance comparisons to obtain a normalized score for each participant controlled for the differences in speed of processing. A similar method was used to compute the individual sensitivity to color proximity. Since some of the processing stages

are common in both numerical and color comparison tasks, we computed the difference between the two RT measures by subtracting the index of sensitivity to color proximity from the index of sensitivity to numerical proximity. The resulting score (dRT) corresponds to a measure of (im)precision of the number magnitude representation, excluding cognitive mechanisms that are not specific to number processing. The higher the dRT score the larger is the overlap between the number magnitudes, and the lower is the precision in comparing symbolic numbers. The individual dRT scores were then entered in regression analyses to examine their influence on task-related brain activity changes.

fMRI ACQUISITION

Blood oxygen level-dependent (BOLD) functional images were acquired in a 1.5 T MRI unit (Gyroscan, Philips Medical Systems), using a multislice T2*-weighted gradient echo-planar imaging (EPI) sequence [TR (repetition time), 3000 ms; TE (echo time), 50 ms; flip angle, 90°] with 33 axial slices, 3.6 mm slice thickness (isotropic voxel), in the bicommissural orientation. The matrix was 64 × 64 and the field of view was 210 × 210 mm. Structural high-resolution T1-weighted 3D gradient echo images (Fast field echo, TR, 30 ms; TE, 3 ms; flip angle, 30°; slice thickness, 1.5 mm) were also acquired.

The stimuli were backprojected using a MRI compatible projector placed at the rear of the magnet and viewed through a tilted mirror mounted on the head coil (Silent Vision® System, Avotec, Inc., http://www.avotec.org). Foam pads were used to restrict head movements. The fMRI paradigm consisted of three runs of eight alternating epochs of comparison tasks (36 s per epoch) and fixation periods (18 s). Each run comprised the acquisition of 144 volumes and contained 160 trials (20 trials × 4 conditions × 2 blocks of each condition per run). Stimulus onset was synchronized with the acquisition of the first slice. The participants received instructions before each sequence, and were not warned of the alternation between tasks and conditions. In each run, two number comparison tasks (close distances and far distances) and two color comparison tasks (close colors and far colors) were presented in pseudo-random order.

fMRI DATA ANALYSIS

Functional data processing and statistical analyses were carried out using Statistical Parametric Mapping (SPM 2, The Wellcome Department of Imaging Neuroscience, London, UK, http://www.fil.ion.ac.uk/spm), implemented in Matlab (Mathworks Inc., Sherborn MA, USA). The first six volumes of each run were discarded to allow for T1 equilibration. All individual images were than realigned to the first remaining fMRI volume of the corresponding participant to correct for within- and between-run motion, coregistered with the individual anatomical scan, and further spatially normalized using the adult MRI template supplied by the Montreal Neurological Institute (MNI). This procedure resulted in normalized fMRI images with a cubic voxel size (2 × 2 × 2 mm). Next, a spatial smoothing with a Gaussian kernel of 8 mm (full width at half maximum, FWHM) was applied in order to reduce the residual anatomical and functional variability across participants. The means (SD) of head movements

of the children in the x, y, and z plane were 0.1 (0.3), 0.6 (0.8), 1.3 (1.4) mm, respectively.

Condition-related changes in regional brain activity were estimated for each participant by a general linear model (GLM) in which the responses evoked by each condition of interest were modeled by a standard hemodynamic response function. The contrasts of interest were first computed at the individual level to identify the cerebral regions significantly activated by numerical (close numerical distance vs. far numerical distance) and color (close color distance vs. far color distance) distances, each condition relative to the fixation periods used as a general baseline. Brain activation maps for the critical contrast (close numerical distance vs. far numerical distance) − (close color distance vs. far color distance) [for a detailed method, see Henson and Penny (2003)] were then entered into a group-level random-effect analysis using a GLM with either age or dRT scores as the covariates of interest. Significant voxels clusters of activation were identified using a threshold of $P < 0.001$ (uncorrected) and an extent threshold of $P < 0.05$, corrected for multiple comparisons, at the cluster level (less than 0.05 under the false discovery rate at the voxel level; see Genovese et al., 2002). The foci that were significantly activated at a corrected $P < 0.05$ (cluster level) or a corrected $P < 0.05$ (FDR, voxel level) were considered. Next, in all brain areas found with RFX analysis, correlation analyses were performed between the beta weights of the contrast and the age or the dRT scores, in order to find out the key area(s) modulated either by age or by the sensitivity to NDE. In this correlation analysis, we applied a cluster size threshold to each significant region resulting from the group-level random-effect analysis using a sphere of 10 mm around the peak of activation. Then, the individual degree of activity (beta values) for each region was correlated with age and dRT score. Anatomical labels were given on the basis of the classification of the AAL (automated anatomical labeling) atlas (Tzourio-Mazoyer et al., 2002).

RESULTS

BEHAVIORAL DATA
Performance in the two tasks was very accurate (95 and 98% of correct responses in numerical and color tasks, respectively) and fast (numerical task: 620 ms; color task: 505 ms). An ANOVA was conducted on median RTs of each participant with task (numerical or color comparisons), and distance (close or far) as within-subject factors. The effect of task was significant [$F_{(1, 18)} = 61.79$, $\eta_{partial} = 0.78$, $P < 0.001$], indicating that the number comparison task was more time-demanding than the color comparison one. As expected, the far distances yielded faster RTs than close ones [$F_{(1, 18)} = 85.53$, $\eta_{partial} = 0.83$, $P < 0.001$] in both tasks [numerical task: $F_{(1, 18)} = 91.36$, $\eta_{partial} = 0.84$, $P < 0.001$, 567 and 673 ms; color task: $F_{(1, 18)} = 5.61$, $\eta_{partial} = 0.25$, $P = 0.03$, 494 and 516 ms] although the impact of this factor was stronger in numerical than color comparisons as indicated by the Task × Distance interaction [$F_{(1, 18)} = 32.49$, $\eta_{partial} = 0.66$, $P < 0.001$].

RELATION BETWEEN AGE AND dRT SCORES
In line with previous data (Holloway and Ansari, 2009; Mundy and Gilmore, 2009), our measure of individual sensitivity to

numerical proximity (dRT) varied from −0.01 to 0.54 ($M = 0.26$; $SD = 0.17$) and showed a negative correlation with age ($r = −0.49$, $P < 0.05$), reflecting a decreasing receptivity to the NDE during development. Moreover, no correlation was found between dRT scores and any neuropsychological measure (all $Ps > 0.05$).

RELATION BETWEEN AGE AND BRAIN ACTIVATION
As illustrated in **Figure 1**, amongst brain regions activated in the contrast [(close – far numerical distance) – (close – far color distance)], significant negative correlations with age emerged in the left superior frontal gyrus ($r = 0.77$, $P < 0.001$), middle frontal gyrus ($r = −0.71$, $P < 0.01$), and IPS ($r = −0.79$, $P < 0.001$). Right lateralized activation foci that decreased with age were also found in the SMA ($r = −0.78$, $P < 0.001$), inferior frontal gyrus ($r = −0.62$, $P < 0.01$), and middle temporal gyrus ($r = −0.67$, $P < 0.01$). Children did not show increase in brain activation with age (whole brain analysis). The impact of age on brain activation in all these brain areas remained significant even when dRT scores were partialled out (see **Table 1**).

RELATION BETWEEN dRT SCORES AND BRAIN ACTIVATION
Children showed only positive correlations between brain activation level and dRT scores used as a measure of sensitivity to the NDE. As illustrated in **Figure 2**, amongst brain regions activated in the contrast [(close - far numerical distance) - (close - far color distance)], the brain activation level was modulated by dRT scores in the middle temporal gyrus in both hemispheres (left: $r = 0.78, P < 0.001$; right: $r = 0.84, P < 0.001$). In the right brain hemisphere, a positive correlation with dRT scores was observed in the superior parietal lobule ($r = 0.80$, $P < 0.001$), in and around the IPS ($r = 0.76$, $P < 0.001$), in the middle occipital gyrus ($r = 0.85$, $P < 0.001$), and in the cingulum ($r = 0.77$, $P < 0.001$). In all the above regions, similar results were obtained when the effect of age was partialled out[3].

We also controlled for the lack of correlation between brain activity level in all significant regions and various individual measures of working memory (forward and backward visuospatial, verbal, and listening spans) and processing speed, such as mean RTs for numerical and color comparison tasks respectively, or mean RTs across both tasks. None of working memory or processing speed measurements correlated with the brain activity in any of the aforementioned brain regions ($Ps > 0.1$). Furthermore, the correlations with dRT scores remained significant even after controlling for these factors (**Table 2**).

DISCUSSION
The present paper investigated changes in neural activity underlying performance during symbolic number comparison in

---

[3]We also analyzed the correlations between the level of activity in areas resulting from the critical contrast (close numerical vs. far numerical) – (close color vs. far color) and the sensitivity to NDE as typically computed in previous studies (Holloway and Ansari, 2009; Mundy and Gilmore, 2009). It appeared that the activity in all regions positively correlated with this "pure" measure of numerical distance effect even when controlled for age (r from 0.23 to 0.70) but not with the similar measure of color distance effect (r from −0.69 to −0.17).
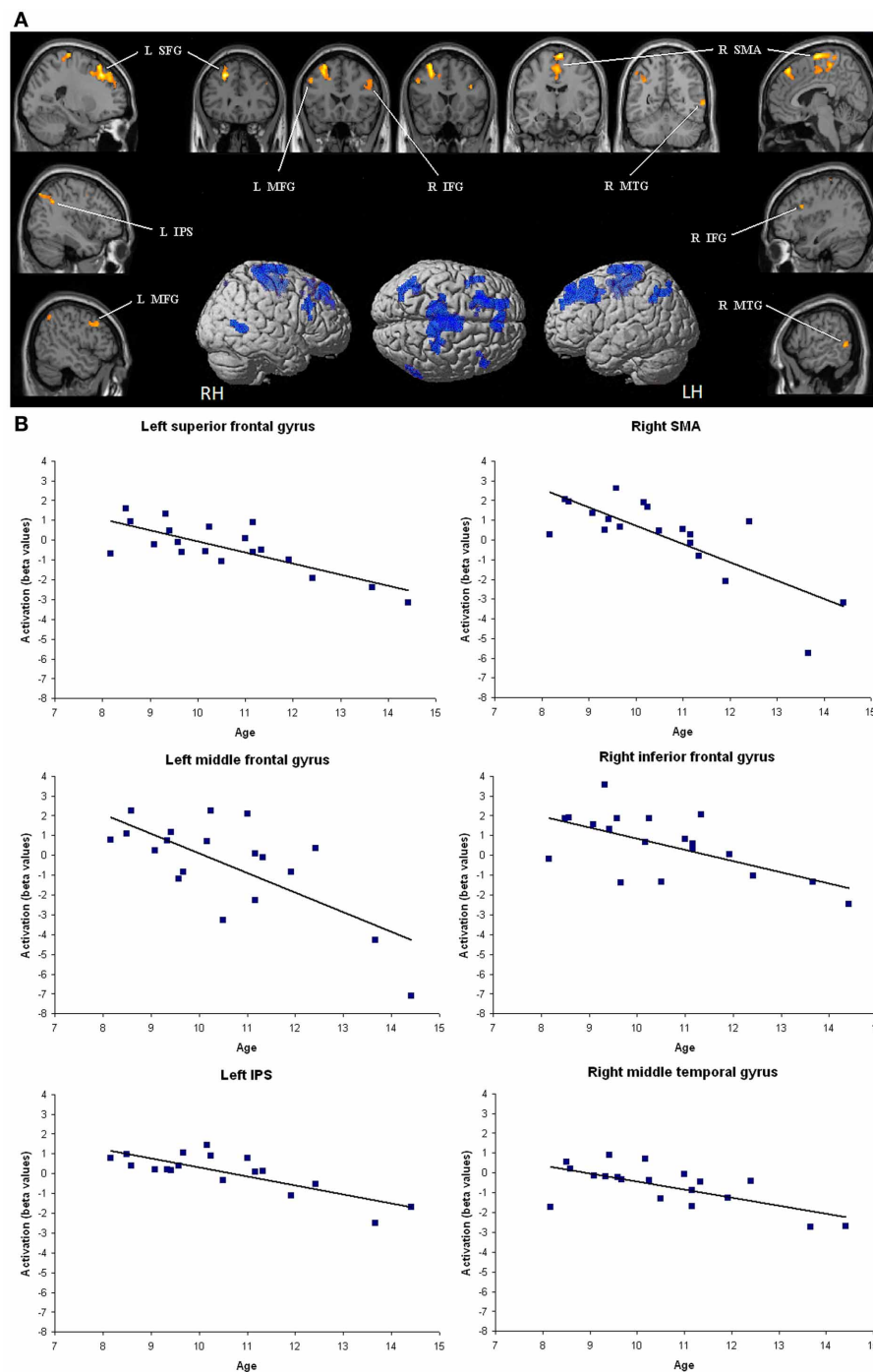
FIGURE 1 | Brain areas that showed significant decreases in activation with age. (A) Brain regions ($P < 0.05$ corrected for multiple comparisons at the cluster level) are superimposed on sagittal and coronal sections of an individual normalized brain MRI. Surface rendering of significant areas of activation that are negatively correlated with age are superimposed on a standard MRI brain for reference. IFG, inferior frontal gyrus; IPS, intraparietal sulcus; MFG, middle frontal gyrus; MTG, middle temporal gyrus; SFG, superior frontal gyrus; SMA, supplementary motor area. **(B)** Bar charts depict decreases in activation (individual beta values) with age in each relevant brain area across children.

children aged 8–14 years. Our results provide pieces of evidence that brain regions showing activity modulation with age were not the same as those affected by the NDE. Here under, the potential role played by each of the regions whose activity was modulated by age or by selective sensitivity to numerical proximity are discussed with regards to past literature.

Amongst areas that showed higher activation for close relative to far numerical distances, significant negative correlation with

Table 1 | Brain areas that showed significant negative correlations with age.

Brain region	No. of voxels	Z_{max}	Talairach coordinates			Correlation	
			x	y	z	r	$r_{partial}$
Left superior frontal gyrus	1181	5.38	−22	30	44	−0.77**	−0.83**
Left middle frontal gyrus	119	3.70	−48	22	36	−0.71*	−0.61*
Left IPS	229	4.00	−38	−52	36	−0.79**	−0.77**
Right SMA	2012	4.94	6	−10	74	−0.78**	−0.79**
Right inferior frontal gyrus	127	3.82	38	18	28	−0.62*	−0.62*
Right middle temporal gyrus	115	4.81	66	−50	4	−0.67*	−0.80**

*Brain regions corresponded to activation peaks, obtained from random-effect analysis using a threshold of P < 0.001, uncorrected, at the voxel level and P < 0.05, corrected for multiple comparisons, at the cluster level. Both simple negative correlations (r) and partial negative correlations (r partial, i.e., corrected for variance explained by dRT scores) are reported. Coordinates are reported in MNI space as given by SPM2 and correspond only approximately to Talairach and Tournoux space (Talairach and Tournoux, 1988). Anatomical labels are based on the AAL (automated anatomical labeling) atlas (Tzourio-Mazoyer et al., 2002). *P < 0.01, **P < 0.001.*

age were found in many frontal regions, including the left middle and superior frontal gyri as well as the right supplementary motor area. This means that, compared to young children, the difference of brain activation between close and far numerical distances in the above areas was weaker in older children. These findings are in line with previous data, which indicated a progressive disengagement of frontal brain regions in symbolic number comparison (Ansari et al., 2005) and calculation (Rivera et al., 2005) with age. Importantly, the difference of brain activation between close and far numerical distances decreased with age not only in frontal regions but also in the left IPS. It is tempting to link these age-related neural changes with the behavioral observations here and elsewhere according to which the size of the NDE decreases during development (Sekuler and Mierkiewicz, 1977; Duncan and McFarland, 1980; Holloway and Ansari, 2009). The question then arises what are the roles of these frontal regions and the left IPS in the NDE. It is worth noting that the brain activity level in these brain regions was not modulated by dRT scores used as individual behavioral indices of selective sensitivity to numerical proximity, which could be considered as indicating imprecision in the number magnitude representation. One could therefore postulate that the frontal brain areas and the left IPS were not dedicated to the number magnitude representation *per se*, but rather were engaged in some cognitive processes related to symbolic number processing, such as the connexion between Arabic digits and the related number magnitudes. As age and experience with symbolic numbers increase, these mechanisms become more and more automatic and require lesser resources, which could be reflected by a weaker engagement of the frontal brain areas and the left IPS during development.

Beyond age-related changes, children showed positive correlations between brain activity level in a series of regions and dRT scores. In several brain areas in the right hemisphere including the middle occipital gyrus, the middle temporal gyrus, the IPS, and the superior parietal lobule, the brain activity level was high in children who were particularly influenced by the numerical proximity between Arabic digits, whereas it was lower in children who were less affected by NDE. This pattern of results is observed not only with dRT scores but also with a "pure" measure of NDE as typically computed in previous studies (Holloway and Ansari, 2009; Mundy and Gilmore, 2009). In our opinion, this demonstrate that the brain regions resulting from our critical contrast are engaged in the manifestation of the NDE, and, more particularly, in number magnitude processing. The role of this right-lateralized occipito-temporo-parietal network has been previously interpreted as reflecting the successive steps of cognitive processes engaged in Arabic number comparison. First, the right middle occipital gyrus and the fusiform gyrus are part of a ventral occipito-temporal pathway specialized for the visual recognition of digits (Cohen and Dehaene, 1996; Pesenti et al., 2000). Second, the IPS, especially in the right hemisphere, is systematically activated when numbers are manipulated, whatever the task and independently of number format (Piazza et al., 2007) or notation (Pinel et al., 2001). In adults, its bilateral recruitment was found to decrease quasi-monotonically as the numerical distance increased, in tight parallel with the behavioral performance (Pinel et al., 2001). The same pattern of change in right IPS activity was also observed in 10-year-old children, albeit to a lesser degree (Ansari et al., 2005). Our data confirm and extend these findings by showing that the brain activity in and around the IPS but also in occipital and temporal regions was modulated by the individual response to numerical proximity between the digits to compare. This indicates some activation of number magnitude in the early stages of number processing. This view is in line with the remarkable demonstration produced by Burr and Ross (2008) who showed that a perceived numerosity is susceptible to adaptation, just like the primary visual properties of a scene, e.g., location, color, or physical size of the stimuli. After a simple 30 s adaptation to a patch containing a large number of spots, an identical following patch seems to have fewer elements. According to the authors, the neurons in the IPS are likely the candidates for the neural substrate of this visual "sense of number." However, as pointed out by Butterworth (2008), earlier neural stages in visual processing could also be involved in the adaptation phenomenon reported by Burr and Ross (2008). This could explain why dRT scores as indices of imprecision in the number magnitude representation also influenced the brain activity in the occipital cortex in the present study.

The present findings favor an involvement of left and right parietal brain areas in number comparison and shed new insight on the specific roles of these brain regions. Whereas the Triple-code model both in its original (Dehaene and Cohen, 1995, 1997) and more recent versions (Dehaene et al., 2003) postulates that left and right IPS areas code the quantity meaning of numbers, past and current research including the present study indicate that the lateralization of number magnitude representation is perhaps more complex. The present observations suggest that the left and
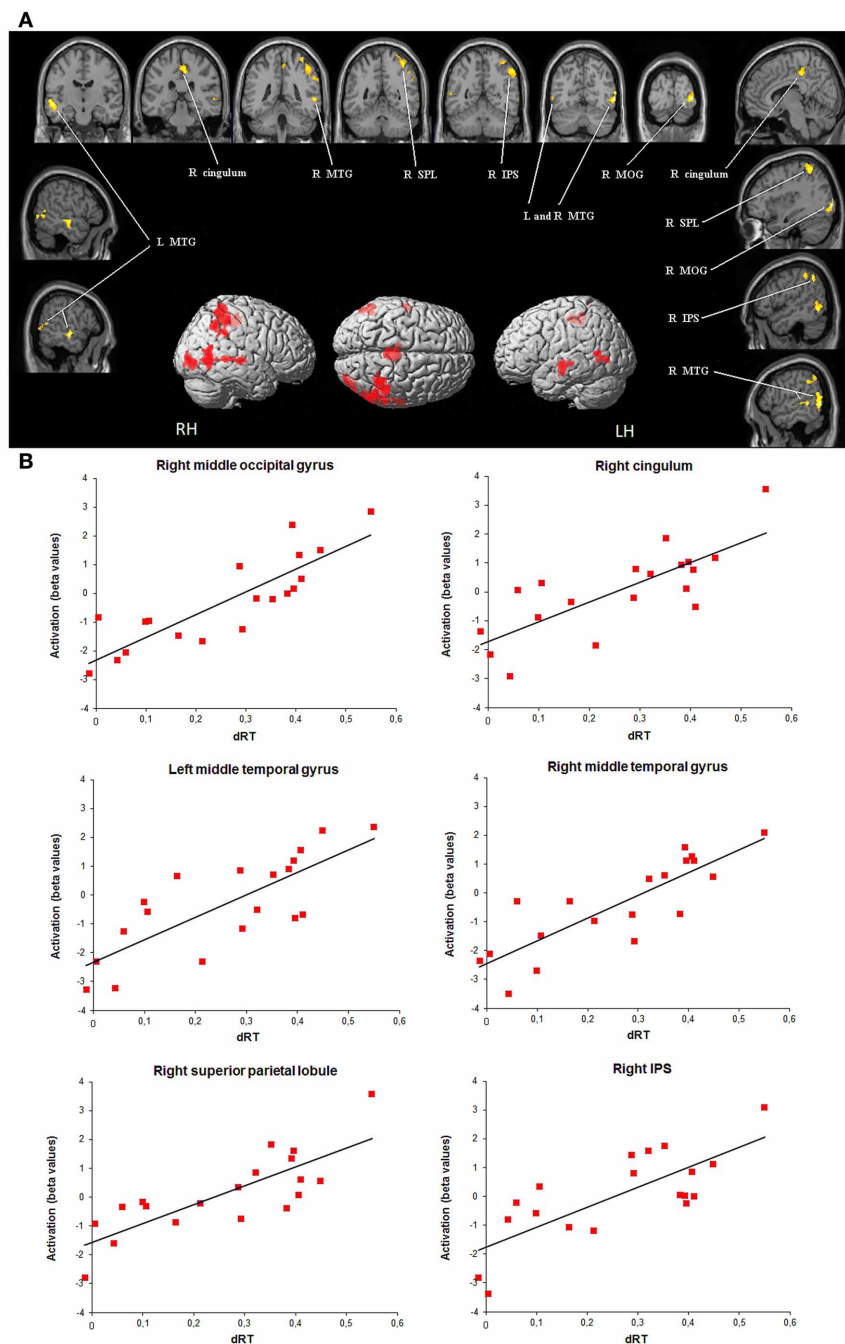
FIGURE 2 | Brain areas that showed significant increases in activation with sensitivity to numerical distance effect (dRT). (A) Brain regions ($P < 0.05$ corrected for multiple comparisons at the cluster level) are superimposed on sagittal and coronal sections of an individual normalized brain MRI. Surface rendering of significant areas of activation that are positively correlated with dRT scores are superimposed on a standard MRI brain for reference. IPS, intraparietal sulcus; MOG, middle occipital gyrus; MTG, middle temporal gyrus; SPL, superior parietal lobule. **(B)** Bar charts depict increases in activation (individual beta values) with dRT scores in each relevant brain area across children. dRT score [adapted from Holloway and Ansari (2009)], individual behavioral index of selective sensitivity to numerical proximity (see text).

right IPS differently contribute to the development of number magnitude processing. The progressive disengagement of regions in the left IPS with increasing age would be related to the refinement of cognitive processes involved but not directly related to number magnitude processing. In contrast, the right IPS would underlie the semantic representation of numbers and its activation would be especially affected by the individual sensitivity to numerical distances between them. This hypothesis accords well

Table 2 | Brain areas that showed significant positive correlations with individual behavioral indexes of selective susceptibility to numerical proximity (dRT scores).

Brain region	No. of voxels	Z_{max}	Talairach coordinates			Correlation			
			x	y	z	r	$r_{partial}$		
							Age	WM[1]	Speed[2]
Left middle temporal gyrus	214	4.41	−56	−14	−8	0.78**	0.73*	0.73*	0.74*
Left middle temporal gyrus	114	3.66	−52	−70	0	0.74**	0.66*	0.68*	0.70*
Right middle temporal gyrus	291	4.50	58	−66	0	0.84**	0.78**	0.77*	0.88**
Right middle temporal gyrus	133	4.37	60	−42	0	0.81**	0.75**	0.83**	0.80**
Right superior parietal lobule	352	4.11	38	−50	58	0.80**	0.76**	0.76*	0.80**
Right lateral bank of IPS—IPS	176	3.84	52	−54	46	0.76**	0.67*	0.77*	0.71*
Right middle occipital gyrus	201	4.79	38	−90	0	0.85**	0.85**	0.82**	0.84**
Right middle cingulum	335	3.90	8	−30	44	0.77**	0.65*	0.70*	0.71*

*Brain regions corresponded to activation peaks, obtained from random-effect analysis using a threshold of $P < 0.001$, uncorrected, at the voxel level and $P < 0.05$, corrected for multiple comparisons, at the cluster level. Both simple positive correlations (r) and partial positive correlations (r partial, i.e., corrected for variance explained by age, working memory, and processing speed measures) are reported. Coordinates are reported in MNI space as given by SPM2 and correspond only approximately to Talairach and Tournoux space (Talairach and Tournoux, 1988). Anatomical labels are based on the AAL (automated anatomical labeling) atlas (Tzourio-Mazoyer et al., 2002). *P < 0.01, **P < 0.001. [1] Working memory measures (forward and backward visuospatial, verbal, and listening spans). [2] Processing speed measures (mean RTs for numerical or color comparisons, mean RTs across tasks).*

with previous reports of a right-hemispheric advantage during number comparison (Dehaene et al., 1996; Chochon et al., 1999; Pinel et al., 2001) as well as other tasks requiring an abstraction of numerical information (Rosselli and Ardila, 1989; Langdon and Warrington, 1997). More direct evidence favoring a right-lateralized representation of number magnitude was provided by Piazza et al. (2007). Using an adaptation paradigm, the authors reported recovery in the right parietal cortex recruitment when Arabic digits or dot patterns changed, and they concluded that this brain area could share neural populations encoding both non-symbolic and symbolic numbers.

In conclusion, the present data extend previous observations by showing that the left and the right IPS as well as other brain regions could contribute differently to number magnitude processing during child development. Further work is clearly needed, however, to determine the precise role of these brain areas in the acquisition of elaborate numerical and non-numerical knowledge.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: Christophe Mussolin, Marie-Pascale Noël, Mauro Pesenti, Cécile Grandin, and Anne G. De Volder. Performed the experiments: Christophe Mussolin and Anne G. De Volder. Analyzed the data: Christophe Mussolin and Anne G. De Volder. Wrote the paper: Christophe Mussolin, Anne G. De Volder, and Mauro Pesenti.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2013.00663/abstract

REFERENCES

Ansari, D., and Dhital, B. (2006). Age-related changes in the activation of the intraparietal sulcus during nonsymbolic magnitude processing: an event-related functional magnetic resonance imaging study. *J. Cogn. Neurosci.* 18, 1820–1828. doi: 10.1162/jocn.2006.18.11.1820

Ansari, D., Dhital, B., and Siong, S. C. (2006). Parametric effects of numerical distance on the intraparietal sulcus during passive viewing of rapid numerosity changes. *Brain Res.* 1067, 181–188. doi: 10.1016/j.brainres.2005.10.083

Ansari, D., Garcia, N., Lucas, E., Hamon, K., and Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *Neuroreport* 16, 1769–1773. doi: 10.1097/01.wnr.0000183905.23396.f1

Burr, D., and Ross, J. (2008). A visual sense of number. *Curr. Biol.* 18, 425–428. doi: 10.1016/j.cub.2008.02.052

Butterworth, B. (2008). Numerosity perception: how many speckles on the hen. *Curr. Biol.* 18, R388–R389. doi: 10.1016/j.cub.2008.03.014

Cantlon, J., Brannon, E., Carter, E., and Pelphrey, K. (2006). Functional imaging of numerical processing in adults and 4-year-old children. *PLoS Biol.* 4:e125. doi: 10.1371/journal.pbio.0040125

Chochon, F., Cohen, L., van de Moortele, P.-F., and Dehaene, S. (1999). Differential contributions of the left and right inferior parietal lobules to number processing. *J. Cogn. Neurosci. Lett.* 11, 617–630. doi: 10.1162/089892999563689

Cohen, L., and Dehaene, S. (1996). Cerebral networks for number processing: evidence from a case of posterior callosal lesion. *Neurocase* 2, 155–174. doi: 10.1080/13554799608402394

Cohen Kadosh, R., Henik, A., Rubinsten, O., Mohr, H., Dori, H., van de Ven, V., et al. (2005). Are numbers special. The comparison systems of the human brain investigated by fMRI. *Neuropsychologia* 43, 1238–1248. doi: 10.1016/j.neuropsychologia.2004.12.017

Corsi, P. M. (1972). *Human Memory and The Medial Temporal Region of The Brain*. Unpublished thesis, Montreal.

Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verb. Learn. Verb. Behav.* 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6

Dehaene, S., and Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Math. Cogn.* 1, 83–120.

Dehaene, S., and Cohen, L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex* 33, 219–250. doi: 10.1016/S0010-9452(08)70002-9

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239

Dehaene, S., Tzourio, N., Frak, V., Raynaud, L., Cohen, L., Mehler, J., et al. (1996). Cerebral activations during number multiplication and comparison: a PET study. *Neuropsychologia* 34, 1097–1106. doi: 10.1016/0028-3932(96)00027-9

Duncan, E., and McFarland, C. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments: an additive-factors analysis. *Mem. Cognit.* 8, 612–622. doi: 10.3758/BF03213781

Eger, E., Sterzer, P., Russ, M. O., Giraud, A.-L., and Kleinschmidt, A. (2003). A supramodal number representation in human intraparietal cortex. *Neuron* 37, 719–725. doi: 10.1016/S0896-6273(03)00036-9

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15, 870–878. doi: 10.1006/nimg.2001.1037

Girelli, L., Lucangeli, D., and Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *J. Exp. Child Psychol.* 76, 104–122. doi: 10.1006/jecp.2000.2564

Henson, R., and Penny, W. (2003). "ANOVAs and SPM," in *Technical Report* (London: Wellcome Department of Imaging Neuroscience).

Holloway, I. D., and Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. *Dev. Sci.* 11, 644–649. doi: 10.1111/j.1467-7687.2008.00712.x

Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children's mathematics achievement. *J. Exp. Child Psychol.* 103, 17–29. doi: 10.1016/j.jecp.2008.04.001

Langdon, D. W., and Warrington, E. K. (1997). The abstraction of numerical relations: a role for the right hemisphere in arithmetic. *J. Int. Neuropsychol. Soc.* 3, 260–268.

Moyer, R. S., and Landauer, T. K. (1967). The time required for judgements of numerical inequality. *Nature* 215, 1519–1520. doi: 10.1038/2151519a0

Mundy, E., and Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *J. Exp. Child Psychol.* 103, 490–502. doi: 10.1016/j.jecp.2009.02.003

Notebaert, K., Pesenti, M., and Reynvoet, B. (2010). The neural origin of the priming distance effect: distance-dependent recovery of parietal activation using symbolic magnitudes. *Hum. Brain Mapp.* 31, 669–677. doi: 10.1002/hbm.20896

Pesenti, M., Thioux, M., Seron, X., and De Volder, A. (2000). Neuroanatomical substrates of Arabic number processing, numerical comparison, and simple addition: a PET study. *J. Cogn. Neurosci.* 12, 461–479. doi: 10.1162/089892900562273

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal

sulcus. *Neuron* 44, 547–555. doi: 10.1016/j.neuron.2004.10.014

Piazza, M., Pinel, P., Le Bihan, D., and Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron* 53, 293–305. doi: 10.1016/j.neuron.2006.11.022

Pinel, P., Dehaene, S., Riviere, D., and LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *Neuroimage* 14, 1013–1026. doi: 10.1006/nimg.2001.0913

Pinel, P., Le Clec'H, G., van de Moortele, P. F., Naccache, L., LeBihan, D., and Dehaene, S. (1999). Event-related fMRI analysis of the cerebral circuit for number comparison. *Neuroreport* 10, 1473–1479. doi: 10.1097/00001756-199905140-00015

Pinel, P., Piazza, M., Le Bihan, D., and Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron* 41, 983–993. doi: 10.1016/S0896-6273(04)00107-2

Rivera, S. M., Reiss, A. L., Eckert, M. A., and Menon, V. (2005). Developmental changes in mental arithmetic: evidence for increased functional specialization of the left inferior parietal cortex. *Cereb. Cortex* 15, 1779–1790. doi: 10.1093/cercor/bhi055

Rosselli, M., and Ardila, A. (1989). Calculation deficits in patients with right and left hemisphere damage. *Neuropsychologia* 27, 607–617. doi: 10.1016/0028-3932(89)90107-3

Rubinsten, O., Henik, A., Berger, A., and Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *J. Exp. Child Psychol.* 81, 74–92. doi: 10.1006/jecp.2001.2645

Schneider, W., Eschmann, A., and Zuccolotto, A. (2002). *E-prime Reference Guide*. Pittsburgh, PA: Psychology Software Tools Inc.

Sekuler, R., and Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Dev.* 48, 630–633. doi: 10.2307/1128664

Talairach, J., and Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*, New York, NY: Thieme Medical.

Temple, E., and Posner, M. I. (1998). Brain mechanisms of quantity are similar in 5-year-olds and adults. *Proc. Natl. Acad. Sci. U.S.A.* 95, 7836–7841. doi: 10.1073/pnas.95.13.7836

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labelling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978

Verguts, T., and Fias, W. (2004). Representation of number in animals and humans: a neural model. *J. Cogn. Neurosci.* 16, 1493–1504. doi: 10.1162/0898929042568497

Wechsler, D. (1996). *Echelle D'intelligence De Wechsler Pour Enfants: Manuel.* 3rd Edn. Paris: Les Editions du Centre de Psychologie Appliquée.

Understanding less than nothing: children's neural response to negative numbers shifts across age and accuracy

Margaret M. Gullick *[†] and George Wolford

Department of Psychological and Brain Sciences, Dartmouth College, Hanover, NH, USA

*Correspondence:
Margaret M. Gullick, Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208, USA
e-mail: margaret.gullick@ northwestern.edu

[†] Present address:
Margaret M. Gullick, Developmental Cognitive Neuroscience Laboratory, Department of Communication Sciences and Disorders, Northwestern University, Evanston, USA

We examined the brain activity underlying the development of our understanding of negative numbers, which are amounts lacking direct physical counterparts. Children performed a paired comparison task with positive and negative numbers during an fMRI session. As previously shown in adults, both pre-instruction fifth-graders and post-instruction seventh-graders demonstrated typical behavioral and neural distance effects to negative numbers, where response times and parietal and frontal activity increased as comparison distance decreased. We then determined the factors impacting the distance effect in each age group. Behaviorally, the fifth-grader distance effect for negatives was significantly predicted only by positive comparison accuracy, indicating that children who were generally better at working with numbers were better at comparing negatives. In seventh-graders, negative number comparison accuracy significantly predicted their negative number distance effect, indicating that children who were better at working with negative numbers demonstrated a more typical distance effect. Across children, as age increased, the negative number distance effect increased in the bilateral IPS and decreased frontally, indicating a frontoparietal shift consistent with previous numerical development literature. In contrast, as negative comparison task accuracy increased, the parietal distance effect increased in the left IPS and decreased in the right, possibly indicating a change from an approximate understanding of negatives' values to a more exact, precise representation (particularly supported by the left IPS) with increasing expertise. These shifts separately indicate the effects of increasing maturity generally in numeric processing and specifically in negative number understanding.

Keywords: integers, negative numbers, fMRI, development, distance effect

The development of numerical cognition includes a progression from an innate approximate recognition of quantity (Xu and Spelke, 2000; Lipton and Spelke, 2003; Xu et al., 2005; McCrink and Wynn, 2007) to a sensitivity to quantity manipulations and violations (Wynn, 1992; Barth et al., 2005, 2006, 2008; Gilmore and Spelke, 2008), an understanding of precise values and the symbols which represent them (Ansari et al., 2005; Gilmore et al., 2007; Roux et al., 2008; Holloway and Ansari, 2009; Lyons and Ansari, 2009; Cantlon et al., 2009), and eventually the ability to perform symbolic arithmetic (Menon et al., 2000; Rivera et al., 2005) and a knowledge of higher-order mathematics (Anderson et al., 2011). These steps reflect increasing expertise with tangible quantities, or symbols representing such concrete amounts. In contrast, understanding negative numbers involves conceptualizing and manipulating abstract quantities that are worth less than nothing yet have value, requiring acceptance of amounts lacking direct physical counterparts. While we have gained mastery of these numbers as adults, negatives remain difficult (Gullick and Wolford, 2013), and the process of conceptual acquisition is unclear. Learning how we come to understand this concept, and whether the brain regions supporting their understanding are similar to or different from those for easier positive numbers, can

inform our knowledge of how we come to understand similarly difficult abstract ideas in both mathematics and other domains, and may be able to eventually inform educational practice and strategies.

Research has recently begun to investigate the behavioral and neural processes supporting our understanding of negative numbers. Studies with adults have indicated that we may have quantitative representations of negatives similar to those for positives (Fischer, 2003; Ganor-Stern and Tzelgov, 2008; Tsang and Schwartz, 2009; Tzelgov et al., 2009; Varma and Schwartz, 2011), and may draw on neural primary number areas for negatives, even if in a slightly less precise or mature manner (Blair et al., 2012; Chassy and Grodd, 2012; Gullick et al., 2012). However, adults have had years of practice and experience with negative numbers and much time to build these positive-like representations. No work has so far described the brain activity underlying children's use of negatives, either before or soon after school instruction on the topic.

Negative numbers are usually introduced into the mathematics curriculum at some point between fourth (Varma and Schwartz, 2011) and sixth (Education, 2010) grade, after years of instruction on and practice with positive numbers. While limited research

has explored children's processing of positive numbers, even less has aimed to examine their negative number understanding. A few works have qualitatively described children's informal verbal explanations of negative numbers, which demonstrate the counterintuitive nature of these items. Negatives seem to be difficult to learn (Streefland, 1996): upon introduction, students may ignore signs (Vlassis, 2004), may inappropriately apply signs (Davis and Maher, 1993), and may order the negative end of the number line backwards (Widjaja et al., 2011). Borba and Nunes (1999) determined that while children had some ability to use negative numbers after a short introduction to the multiple meanings of the minus sign, their explanations and demarcations were idiosyncratic and most were unable to use manipulable materials in their explanations, preferring only oral descriptions. After two years of instruction, though, Varma and Schwartz (2011) demonstrated that sixth-grade children showed behavioral effects similar to those for adults for negative number comparisons. As such, children may not initially understand negative numbers, but come to use them fluently after some practice and experience. We here specifically investigate the neural systems supporting negative number use, compared to those involved in positive number processing, in pre- and post-instruction children in an effort to better understand the trajectory of the acquisition of negative number knowledge from initial processing strategies to eventual adult expertise.

Most often, investigations of adult negative number processing have examined the presence and direction of the distance effect, a typical positive-number processing result. First described by Moyer and Landauer (1967), the distance effect describes an inverse relationship between comparator distance and reaction times: comparisons involving numbers that are further apart (at a greater distance) are responded to faster and more accurately than those with numbers that are closer together. This effect has been taken to reflect the possible representational overlap between neighboring numbers on a mental number line, as ordering and choosing the greater of two numbers becomes more difficult to resolve if the items are closer together (Van Opstal et al., 2008; Holloway and Ansari, 2010; Holloway et al., 2010). Neurally, the distance effect is reflected in gradations of IPS activity: close-distance numeric comparisons elicit more IPS activity than far comparisons (Pinel et al., 2001; Ansari et al., 2006), possibly reflecting the increased effort and neural activity needed to resolve the more difficult closer comparisons.

The distance effect is conserved in children, though there are developmental changes in scale and localization. Behaviorally, while overall response times are slowed, children still demonstrate faster and more accurate responses to farther than closer comparisons (Holloway and Ansari, 2008). The size of the effect also decreases with age, as children show a greater distance effect than adults (Holloway and Ansari, 2008), and may also be related to mathematics achievement, with the effect becoming smaller and less dramatic with increasing math skill (De Smedt et al., 2009; Holloway and Ansari, 2009; but see Schneider et al., 2009).

Like adults, children may also demonstrate a significant effect of distance neurally, but the pattern of responses differs by age. The IPS demonstrates a distance effect to non-symbolic numbers from at least age four (in an adaptation paradigm; see Cantlon

et al., 2006) or age six or seven (in a comparison paradigm; see Ansari and Dhital, 2006; Cantlon et al., 2009). Comparison effects for symbolic numbers, though, may not be IPS-based. Instead, children (ages 8–12) may show only a frontal distance effect, particularly in the dorsolateral prefrontal cortex and inferior frontal gyrus. The IPS does demonstrate significant activity during comparisons, but is not distance modulated (Ansari et al., 2005). Temple and Posner (1998) did note a parietal distance effect for both symbolic and non-symbolic comparisons in five-year-olds using event-related potentials, indicating that parietal number areas may be responsive to symbolic numbers even from a young age, but the spatial resolution of ERPs makes source localization difficult. At some point, the distance effect may shift to be both frontal and parietal, then to the adult parietal-only effect, but the trajectory of these changes has not been fully described. Generally, though, these shifts are consistent with the developmental finding wherein activity related to mathematics processing shifts posteriorly with age (Rivera et al., 2005).

Behaviorally, several studies have tested adult responses to negative numbers, most often in paired comparison paradigms. Negative pairs have consistently demonstrated a typical distance effect in both simultaneous (Varma et al., 2007; Tzelgov et al., 2009; Varma and Schwartz, 2011) and sequential (Ganor-Stern et al., 2010) single-digit negative number paired comparisons. Recently, Gullick, Wolford, and Temple (Gullick et al., 2012) also tested the adult brain activity supporting paired comparisons with positive and with negative numbers. First, across comparison distances, negative numbers showed increased parietal-lobe activity, including in the IPS, relative to positive pairs, along with increased caudate and decreased frontal-lobe activity. This overall parietal increase may be due to differences in difficulty between the signs (Gobel et al., 2004; Blair et al., 2012; Chassy and Grodd, 2012). Importantly, though, Gullick et al. (2012) found that negatives also showed typical distance-modulated responses both behaviorally and neurally, including in the IPS, which was also greater for negative than for positive pairs in each case; positive comparisons again demonstrated a stronger effect of distance frontally than did negatives. This difference indicates that while processing abstract negative numbers strongly engages primary number areas, it may less draw on frontal secondary regions than concrete positive numbers. As a greater distance effect has been taken to indicate a less mature representation of number (De Smedt et al., 2009; Holloway and Ansari, 2009), the representation of negative numbers was proposed to be less precise than that for positives, leading to this more dramatic distance effect.

Adults may thus understand negative numbers as individual quantities arranged along the leftward end of a bidirectional number line. These representations seem to be supported by similar activity in the same quantity-sensitive regions as positive numbers. This mature usage, though, may stem from years of experience and practice with negatives. How do children, who have little or no formal experience with these concepts, respond to negative numbers?

In summary, numeric negativity appears to be a difficult concept to acquire. Children may be able to use negatives in some limited situations even before instruction, but such knowledge is typically limited to informal situations (Mukhopadhyay

et al., 1990) and is unstable (Borba and Nunes, 1999). After some instruction, children may demonstrate a typical behavioral distance effect to negative comparisons, (Varma and Schwartz, 2011), but whether children also show a typical neural distance effect for negatives, or after what amount of practice such an effect may appear, is not known.

We here aimed to examine whether children showed a neural distance effect for negative number comparisons similar to that for positive numbers. In our area (New Hampshire and Vermont), negative numbers are formally introduced in the sixth grade (2010). As such, fifth-graders were used as a pre-instruction group, and seventh-graders as a post-instruction cohort. Fifth-graders were not expected to be completely naïve to negatives, so should be able to perform simple comparisons but not able to use negatives in arithmetic or more complex situations. Seventh-graders, who should have at least one year of formal experience with negatives, were expected to show a greater proficiency with negatives, including in comparisons.

A distance effect for positive number comparisons was expected in certain neural locations within each age group. In line with Ansari et al. (2005), the ten- to eleven-year-old fifth-graders were expected to demonstrate a frontal distance effect for positive-number comparisons. The 12- to 13-year-old seventh-graders were at or beyond the upper limit of Ansari et al. (2005), and could show both frontal and parietal distance effects for positive number comparisons. The presence, direction, and location of a negative number comparison distance effect was then explored within each age group. This method allows determination of whether effects "match" across signs, even as the effect location shifts between age groups, and thus whether negative numbers are processed using the same neural mechanisms as positives at each instructional stage (e.g., whether negative comparisons evoke the same frontal or parietal distance effect as positive pairs).

METHODS

PARTICIPANTS

Participants were 16 (6F) fifth-graders, ages 9;11–11;9 (mean = 10;8 years), and 15 (5F) seventh-graders, ages 11;9–13;5 (mean = 12;8 years; see **Table 1** for demographic information). All were right-handed, as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), with no history of learning disorder or neurological damage. Nine additional participations were excluded, 4 due to artifact from braces, 2 due to response recording problems, 2 due to excessive movement, and 1 due to use of drugs affecting white matter integrity. Participants were paid $25 and given several small prizes.

STIMULI

Stimuli were the same as used by Gullick et al. (2012) with adults. Briefly, stimuli used pairs of numbers from −20 to 20, excluding zero. Comparison pairs were created in three main sign categories (20 positive, 20 negative, and 60 mixed comparisons), for a total of 80 unique pairs (see **Table 2** for example pairs from each included category; we here focus on only positive and only negative comparison pairs, and so mixed pairs are not further discussed). Half the comparisons in each sign type were closer in

Table 1 | Demographic characteristic of each age group.

	Fifth-graders	Seventh-graders	
N	16	15	
Age	10;8 (7.66 months)	12;8 (5.89 months)	$p < 0.001$
Age range	9;11–11;9	11;9–13;5	
KTEA-II Math Concepts and Applications standard score	114.06 (17.48)	120.73 (9.71)	$p > 0.2$
Integer Knowledge Test (out of 62)	36.81 (7.34)	51 (7.55)	$p < 0.001$
Stroop interference standard score	52 (6.53) (14 participants)	50.47 (9.01)	$p > 0.6$

Table 2 | Example stimuli.

Distance groupings:	Sign category	
	Positive	Negative
Closer (distances 1–8)	3 <> 5	−3 <> −5
Farther (distances 12–19)	3 <> 16	−3 <> −16

distance (between 1 and 8) and used two single-digit comparators, while half were farther apart (between 12 and 19) and used one single- and one double-digit comparator. Positive and negative number comparisons used the same digits, but different signs.

Half the comparisons involved two presented digits, and half used thermometers. While fifth-graders were not expected to have formal experience with negatives as abstract digits, they could recognize negatives as representing very cold temperatures, given their geographic location. As such, comparisons were presented in half the runs as digits, and in half as temperatures on a thermometer. Thermometers were created using a blank canonical shape, with unlabeled side tics and red filling up to half-height. Temperature was labeled in red to the left of the thermometer by the middle tic. These thermometers were meant to invoke and reinforce the context of a temperature comparison, but not to test thermometer-reading skill. To keep participants from ignoring the numbers presented and simply visually comparing the amount of "red stuff" (mercury) in each thermometer, filling height and digit position were kept constant, making it impossible to base the comparison off area or relative vertical number-line position. As this comparison, and indeed the range of temperatures presented, was possible in either Celsius or Fahrenheit, no specific scale was given for the thermometers: participants were simply instructed to think of the numbers as temperatures and choose the warmer (or colder) temperature. We here focus on responses collapsed across presentation format.

Baseline control trials presenting a blank screen were also used.

PROCEDURE

After obtaining informed consent, participants completed a survey testing their knowledge of signed number usage and operations (see Appendix), which included questions requiring participants to order numbers, choose the larger number, and perform

simple arithmetic with and complete word problems involving negative numbers. Participants were also given the Math Concepts and Applications subtest of the Kaufmann Test of Educational Achievement-II (KTEA-II; Kaufman and Kaufman, 2004), which tests the ability to apply mathematics knowledge to solve problems. A color-word Stroop test standardized for children (Golden et al., 2003) was also administered to all the seventh-graders, and 14 of the 16 fifth-graders. The Stroop test was administered in the same testing session as the fMRI scan for 20 of the included participants (9 5th graders and 13 7th graders), and in a separate second session between 2.5 and 4 months later for the remaining participants. Two fifth-graders could not return for a second testing session, and so Stroop scores are reported for only the 14 available participants. Scores were normed to age of test administration.

After a short practice session, fMRI data was acquired in four event-related functional runs. One half of the experimental session asked participants to choose the larger number (or warmer temperature), the other half asked participants to choose the smaller number (or colder temperature). Question and format order were counterbalanced across participants. Run lists consisted of 100 experimental trials, using one instance (left- or right-greater) of each unique comparison. Thirty-two baseline control trials were also included in each run (including three at the end of each run) at jittered intervals. Each run was ~5.5 min in length. Stimulus pairs were presented in pseudorandom order. Each unique comparison was thus presented in eight variations (within each context, in each question version, with the greater number on the left vs. right).

In each comparison trial, one item was presented on the left side of the screen, and one on the right, separated by a "<>" symbol. Digit comparison pairs were presented for 1.5 s, followed by a 1 s blank screen, and thermometer comparison pairs for 2 s, followed by a 500 ms blank screen (see **Figure 1**). Participants could respond at any point within the display time, but were encouraged to respond quickly and accurately. Behavioral piloting determined that these presentation periods gave participants ample time to answer on each trial while continuing to encourage speeded responses; longer presentation times for digits may have resulted in decreased task attention and less pressure to respond immediately. Stimulus presentation, trial timing, and response recording was achieved using E-Prime presentation software (Psychological Software Tools, Pittsburgh, PA). Stimulus pairs were presented using a Panasonic DT-4000U DLP projector,

and each functional run was synchronized with the onset of the first trial to ensure accuracy of event timing. Response times and accuracy were measured using fiber optic button press boxes (Cedrus Lumina response pads; San Pedro, CA).

DATA ACQUISITION

Functional images were acquired in a 3T Philips Achieva Intera MRI scanner at the Dartmouth Brain Imaging Center. In each of the four functional imaging runs, we acquired 132 whole-brain T2*-weighted echoplanar images (EPI). 41-slice whole-brain EPI image volumes were acquired using Philips interleaved sequence maximizing the distance between neighboring slices; here, slices were acquired in intervals of 6. The following parameters were used for acquisition: slice thickness = 3 mm, no skip; repetition time (TR) = 2.5 s; echo time (TE) = 35 ms; flip angle = 90°; matrix = 80 × 80; field of view (FOV) = 240 mm; transverse plane. Two additional volumes were discarded at the beginning of each run to allow for equilibrium effects. In addition, a high resolution, magnetization-prepared rapid-acquisition gradient echo (MPRAGE) image was acquired at the end of the session, but was not used for analysis.

ANALYSIS
Behavioral analyses

Behavioral data were analyzed to determine response accuracy and ensure task attention, as well as examine the presence and direction of a distance effect, in each age group. Comparison sign types (positive, negative) were first examined for overall response time and accuracy differences. The presence and direction of any distance effects in each sign type was examined through a series of linear regressions, performed on each participant's correct response data, to determine whether response times were significantly predicted by comparison distance. The unstandardized beta coefficient for the distance regressor for each individual was then extracted, and compared across sign types using ANOVAs, with a statistical threshold of $p < 0.05$.

fMRI analyses

All functional data were examined for artifact by creating signal to noise maps in MATLAB (version 7.7.0 R2008b; The MathWorks, Inc., Natick, MA) with a modified script available at http://dbic.dartmouth.edu/wiki/index.php/Noise_Detection. fMRI data were processed using SPM8 (Welcome Department of Cognitive Neurology, London, UK, http://www.fil.ion.ucl.ac.uk/spm). Preprocessing steps for each participant included the following steps. *Reorientation:* The center of each functional image was reoriented such that the origin was at the midsagittal anterior commissure. *Slice Timing Correction:* Differences in image acquisition time between slices were corrected using the first slice as reference using SPM8's Fourier phase shift interpolation. *Realignment:* Head motion was realigned to the mean image using the least-squares approach and a 6-parameter rigid-body spatial transformation. Estimation was performed at 0.9 quality, 4 mm separation, 6 mm FWHM smoothing kernel, using second degree B-Spline interpolation. Reslicing was performed using fourth degree B-Spline interpolation. The realignment parameters were examined for excessive motion (defined as >1 mm
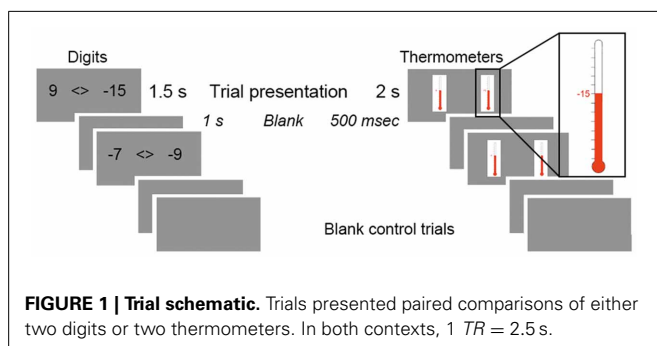
FIGURE 1 | Trial schematic. Trials presented paired comparisons of either two digits or two thermometers. In both contexts, 1 TR = 2.5 s.

motion in any direction). Two participants were excluded for head position drift exceeding 4 mm and multiple occurrences of movement spikes exceeding 2 mm. *Smoothing:* Images were smoothed using a 6 mm FWHM Gaussian kernel.

First-level individual statistics. All runs from each individual were analyzed together using a mass-univariate approach based on the general linear model. Two factors were modeled: comparison context (digits, thermometers), and comparison pair sign type (positive, negative, mixed polarity sensitive, mixed polarity insensitive), along with baseline control trials and the six realignment parameters from motion correction as parameters of no interest for each run. Control trials were thus modeled explicitly (see Poline et al., 2003). Error trials were included as a condition in individual analysis. A first-order parametric modulator was also included for each comparison pair sign type, marking the distance between comparators (1–19) on each trial. A high-pass filter of 128 s was used to remove slow signal drift. Summary contrast maps were created for each individual to take to second-level group analysis. Based on the specific planned group-level tests, normalized contrasts of each experimental condition versus the modeled baseline control were performed (e.g., positive > baseline, negative > baseline, etc.). The contrast for each stimulus class' parametric modulator was also created (e.g., positive distance effect > baseline). Contrasts using this modulator determined areas of the brain whose activity changes linearly in accordance with changes in distance. Mask images for each individual were examined to ensure full brain coverage. This analysis is thus the same as that used with adults (see Gullick et al., 2012).

Data from all participants was normalized to the standard (adult) SPM8 EPI template using a trilinear interpolation, writing 3 mm^3 voxels. While pediatric templates could be used, several factors argued against their implementation in this study. First, the children included here were easily over 7 years of age, considered to be the point when the brain reaches ∼95% of its adult size (see Caviness et al., 1996): fifth- and seventh-graders can thus be considered to have nearly volumetrically-mature brains, though cortical thickness and mylenation continue to develop and change. Further, previous work has demonstrated that the differences between activity localization in 7- and 8-year-old children and adults may be negligible, given the voxel sizes and smoothing kernels used in conventional fMRI analysis (Burgund et al., 2002; Kang et al., 2003). Last, some analyses compare activation across age groups. Such contrasts are best conducted on data that has all been normalized to the same space, as warping to different templates may result in systematic misregistrations and activity mislocalizations (see page 68 in Poldrack et al., 2011).

Second-level group statistics. A randomized effects model was used for group analysis. All analyses were performed within a mask of frontal (IFG, MFG, SFG) and parietal (IPL, SPL) cortex (see **Figure 2**). This mask was created by combining the WFU pickatlas definitions of these regions, as implemented through the SPM8 toolbox. Comparisons between stimulus classes were first performed to examine differences in brain activity based on

FIGURE 2 | Parietofrontal mask for fMRI analyses. All fMRI analyses were performed within a mask of the frontal (IFG, MFG, SFG) and parietal (IPL, SPL) lobes.
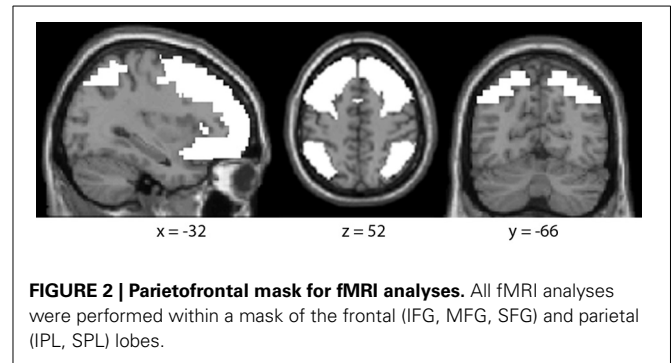
sign type. These results are reported at thresholds of peak voxel level $p < 0.005$ (uncorrected), cluster $p < 0.05$ (FDR corrected), cluster size $k > 30$.

Comparisons were then made examining areas that showed increasing activity in response to decreasing comparison distance (i.e., a typical distance effect), or in response to increasing distance (a reversed distance effect), as found with the behavioral distance effect analyses, for each comparison sign type. These analyses examined brain activity associated with the parametric modulator of distance for each sign type. The parametric modulator function looks for areas of the brain that show a significant relationship between the modulator (here, distance) and brain activity, beyond any activity variance accounted for by the main effects (here, comparison sign type). Distance was thus used as a continuous measure, not binned into distance categories. As such, this represents a further and more stringent analysis, as some variance has already been accounted for by sign type, and thus results are not as strongly significant as with the main effects. Given this difference, the results for activity related to the modulator of distance are reported at thresholds of peak voxel level $p < 0.05$, cluster size $k > 10$.

Co-ordinates are MNI using ICBM152. Anatomical regions were assigned by a combination of xjview (Cui et al., 2011), visual inspection, and Talairach daemon after transformation to Talairach space (Lancaster et al., 1997, 2000; Brett, 2006). The anatomical region listed is for the peak voxel location. In all cases, IPS activity was confirmed by hand if the analyses demonstrated significant activity peaks in the inferior or superior parietal lobules.

RESULTS
BEHAVIORAL TESTING
Independent samples t-tests demonstrated that seventh-graders were significantly older than fifth-graders [$t_{(29)} = 9.67$, $p < 0.001$]. Neither KTEA-II standard scores [$t_{(29)} = 1.301$, $p = 0.204$] nor Stroop interference standard scores [$t_{(27)} = 0.522$, $p = 0.606$] were significantly different between age groups, but seventh-graders performed significantly better than fifth-graders on the Integer Knowledge Test [$t_{(29)} = 5.304$, $p < 0.001$] (see **Table 1** for group scores and means): no fifth-graders were able to perform multiplication or division with negative numbers, and all but three could not perform addition or subtraction, while all seventh-graders were at least able to perform addition and subtraction.

EFFECTS OF SIGN AND AGE

A 2 (sign) × 2 (age group) Repeated Measures ANOVA was first performed on accuracy data. Between subjects, there was no significant effect of age [$F_{(1, 29)} = 1.71$, $p > 0.2$, $MSE = 0.019$], indicating that fifth-grader (mean = 84.8%) and seventh-grader (mean = 88.3%) performance was similar on the task. Within subjects, there was a significant effect of sign, $F_{(1, 29)} = 38.238$, $p < 0.001$, $MSE = 0.130$, $\eta_p^2 = 0.569$, where responses to positive number comparisons (mean = 91.1%) were more accurate than to negative number comparisons (mean = 81.9%), but no significant sign × group interaction was found [$F_{(1, 29)} = 3.166$, $p = 0.086$, $MSE = 0.011$].

A 2 × 2 Repeated Measures ANOVA was then performed on response time data. Between subjects, there was a significant main effect of age, $F_{(1, 29)} = 4.777$, $MSE = 221550$, $p = 0.037$, $\eta_p^2 = 0.141$. Within subjects, there was a significant main effect of sign, $F_{(1, 29)} = 98.822$, $MSE = 272424$, $p < 0.001$, $\eta_p^2 = 0.773$, and a significant interaction between sign and age, $F_{(1, 29)} = 8.247$, $p = 0.008$, $\eta_p^2 = 0.221$. Seventh-grader responses (mean = 967.75 ms) were faster than fifth [mean = 1087.37 ms; $t_{(29)} = 2.186$, $p = 0.037$], and positive number comparison responses (mean = 961.24 ms) were faster than negative [mean = 1093.88; $t_{(30)} = -9.009$, $p < 0.001$]. Fifth- and seventh-grader response times did not differ for positive number comparisons [$t_{(29)} = 1.527$, $p > 0.1$], but seventh-grader responses were significantly faster than fifth for negative number comparisons [$t_{(30)} = 2.664$, $p = 0.012$].

DISTANCE EFFECTS

The presence and direction of a distance effect in participant response times was assessed using linear regressions. Response times for each comparison pair type were used as the dependent variable, and numeric comparison distance as the independent, to determine whether response time was significantly predicted by distance. Unstandardized beta coefficients for the distance regressor were extracted for each participant, once for the positive pair comparisons and once for the negative. The beta coefficients reflect the predicted change in response time (in milliseconds) for a one-unit change in distance; negative beta coefficients indicate a typical distance effect, as response times should decrease as distance increases, and positive beta coefficients a reversed distance effect. This analysis thus allows examination of the effect of distance on response time, but also removes the effect of overall response time differences between the sign categories. This analysis is the same as that used with adults (see Gullick et al., 2012). These beta-weights were then entered into one-sample t-tests to determine whether they were significantly different from zero, and thus showed a significant effect of distance on response times across individuals.

Fifth-graders demonstrated a typical distance effect for negative number comparisons [mean = −6.05, $t_{(15)} = -3.565$, $p = 0.003$], but no significant effect for positive number comparisons [mean = −1.72, $t_{(15)} = -1.253$, $p = 0.229$]. Seventh-graders demonstrated a typical distance effect for negative number comparisons [mean = −5.36, $t_{(14)} = -5.238$, $p < 0.001$], but no significant effect for positive number comparisons [mean = −2.19, $t_{(14)} = -1.516$, $p = 0.152$] (see **Figure 3**). These beta coefficients
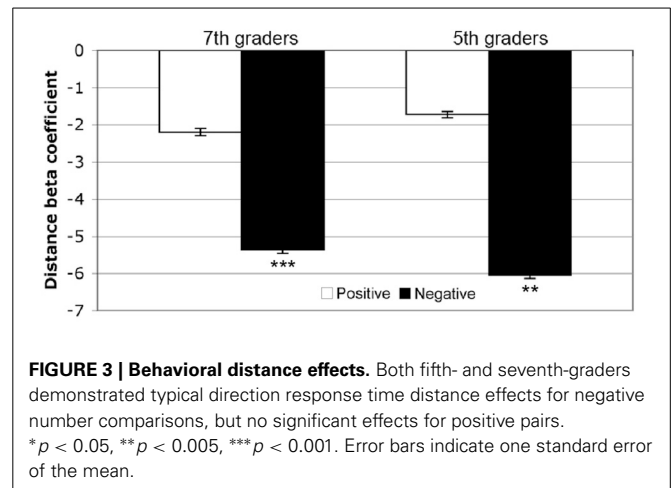
FIGURE 3 | Behavioral distance effects. Both fifth- and seventh-graders demonstrated typical direction response time distance effects for negative number comparisons, but no significant effects for positive pairs. *$p < 0.05$, **$p < 0.005$, ***$p < 0.001$. Error bars indicate one standard error of the mean.

were then compared between age groups using a 2 (sign) × 2 (age group) Repeated Measures ANOVA. Between subjects, there was no significant main effect of age ($F < 1$). There was a significant main effect of sign, $F_{(1, 29)} = 10.445$, $MSE = 217$, $p = 0.003$, $\eta_p^2 = 0.265$, but no interaction between sign and age [$F < 1$]. Negative comparison pairs thus demonstrated a stronger distance effect than positive in each age group.

Positive number comparisons thus did not demonstrate a significant distance effect in these context-collapsed analyses, though negative number comparisons did. To further investigate this situation, we conducted separate analyses for each sign within each presentation context. Fifth-graders demonstrated a marginally significant typical distance effect for digit-context positive number comparisons [mean = −5.94, $t_{(15)} = -2.063$, $p = 0.057$], but not for negative numbers [mean = −2.42, $t_{(15)} = -1.044$, $p > 0.3$]. However, temperature-context comparisons showed a significant distance effect for negative number comparisons [mean = −11.11, $t_{(15)} = -5.406$, $p < 0.001$] but not positive [mean = 2.94, $t_{(15)} = 1.700$, $p > 0.1$]. Seventh-graders demonstrated significant distance effects for both signs in the digit context [positive number pairs: mean = −4.62, $t_{(14)} = -3.158$, $p = 0.007$; negative number pairs: mean = −3.59, $t_{(14)} = -2.585$, $p = 0.022$]. Similarly to the younger subjects, temperature-context comparisons showed a significant distance effect for negative number comparisons [mean = −7.79, $t_{(14)} = -3.886$, $p = 0.002$] but not positive [mean = 0.64, $t_{(14)} = 0.316$, $p > 0.7$]. As such, positive numbers always demonstrated a typical distance effect in digit-format presentations, but not when presented as thermometers, for both age groups. Negative numbers showed a significant distance effect when presented as thermometers in both groups, but a significant digit-format effect was found only for seventh graders. Despite these context-dependent outcomes, there were insufficient trials to separately analyze other potential digit- vs. thermometer-format trial effects. Thus, all further analyses collapse across presentation context, combining digit- and thermometer-format pairs within each sign.

PREDICTING THE SIZE OF THE DISTANCE EFFECT

We then investigated the factors predicting the size of the distance effect (beta coefficient) for negative number comparisons

across individuals within each age group using stepwise linear regressions. In each regression, the negative number comparison beta coefficient was entered as the dependent variable, and negative comparison accuracy, negative response time, positive comparison accuracy, positive response time, positive comparison beta coefficient, age (in months), KTEA-II standardized score, Integer Knowledge Test score, and Stroop interference score (where applicable) were included as independent variables.

In fifth-graders, only positive comparison accuracy significantly predicted the size of the negative number comparison distance effect, $t_{(14)} = -3.45$, $p = 0.004$. In seventh-graders, the size of the negative number comparison distance effect was predicted first by negative comparison accuracy, $t_{(14)} = -3.24$, $p = 0.007$, then additionally by Stroop interference score, $t_{(14)} = 3.28$, $p = 0.007$ (see **Figure 4**).

fMRI RESULTS

EFFECTS OF SIGN AND AGE

As with the behavioral analyses, data was first compared between sign types, across comparison distances, to determine areas differentially involved in processing positive and negative numbers within each age group (see **Table 3**, **Figure 5**). In fifth-graders, positive number comparisons showed greater activity than negative comparisons in the bilateral parietal lobe, including the IPS, as well as the bilateral inferior and superior frontal gyri; negative comparisons evoked more activity than positive in one cluster in the right inferior parietal lobule, though not in the IPS. In seventh-graders, negative comparisons evoked more activity than positive in one cluster in the left inferior frontal gyrus. No further differences were seen.

DISTANCE EFFECTS

fMRI data was then analyzed to determine brain areas showing a linear increase in activity as comparison distance increased, or decreased, within each sign category. Significant clusters were identified at a height threshold of $p < 0.05$ and cluster size of 10 voxels. In fifth-graders, positive comparisons demonstrated a typical distance effect both frontally (including the left inferior

frontal and right precentral gyri) and parietally (including the left IPS). Negative comparisons also showed a typical distance effect across the frontal and parietal lobes, though in only the right IPS. This effect was greater for positive comparisons in the parietal lobe, including the right IPS, but was greater for negative

Table 3 | Positive versus negative comparisons, within each age group.

Location of peak voxel	MNI coordinates			Cluster size	Peak t
	x	y	z		
FIFTH GRADERS POSITIVE > NEGATIVE NUMBERS					
R inferior frontal gyrus	12	38	−20	89	5.92
R superior frontal gyrus	18	38	58	54	5.7
L superior, middle frontal gyrus	−21	38	58	148	5.61
L inferior frontal gyrus	−60	29	−2	121	5.6
L superior frontal gyrus	−12	−10	67	122	5.46
R inferior parietal lobule	42	−70	46	58	5.3
L superior frontal gyrus	−18	44	−17	101	4.97
R middle frontal gyrus	48	56	10	54	4.69
R inferior, middle frontal gyrus	60	14	31	214	4.51
R superior frontal gyrus	−9	62	1	68	4.01
L inferior parietal lobule	−42	−64	40	49	3.76
FIFTH GRADERS NEGATIVE > POSITIVE NUMBERS					
R inferior parietal lobule	42	−34	28	49	5.2
SEVENTH GRADERS POSITIVE > NEGATIVE NUMBERS					
None					
SEVENTH GRADERS NEGATIVE > POSITIVE NUMBERS					
L inferior frontal gyrus	−33	23	−11	47	4.01
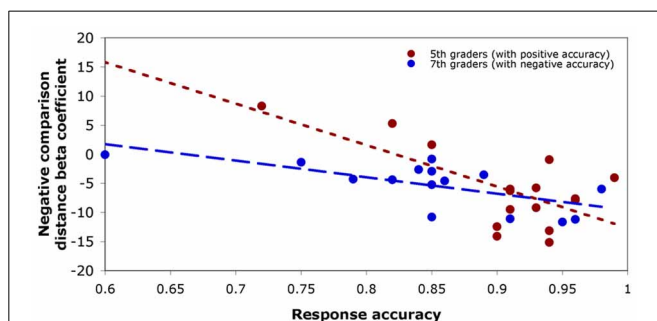
FIGURE 4 | Predicting the size of the negative number distance effect. Fifth-graders (red) demonstrated a significant negative relationship between positive comparison accuracy and negative comparison distance effect. Seventh-graders (blue) demonstrated a significant negative relationship between negative comparison accuracy and negative comparison distance effect. In each case, participants with higher comparison accuracies showed a more negative (typical) distance effect.
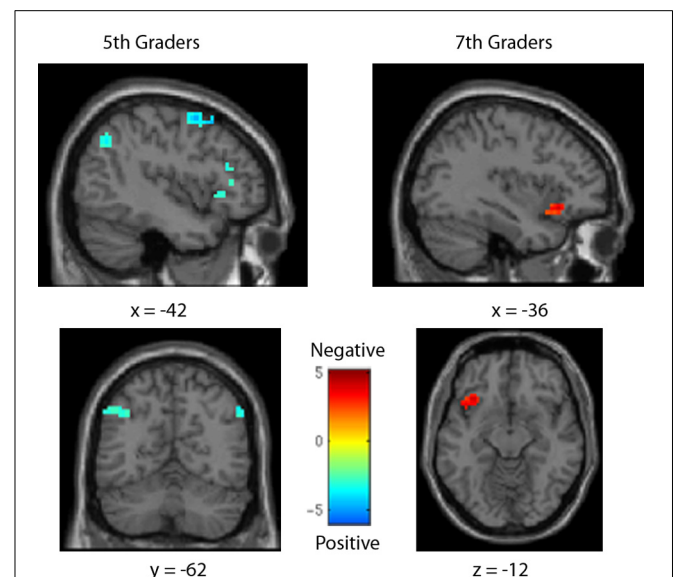
FIGURE 5 | Brain activity for positive vs. negative number comparisons in each age group. For fifth-graders, positive number comparisons (cool colors) showed greater activity in the left IPS. For seventh graders, negative number comparisons (warm colors) demonstrated greater activity in only the left inferior frontal gyrus. *All fMRI figures show left-hemisphere activity on the left side. All figure colorbars indicate t-test contrast values.*

Table 4 | Fifth-grader neural distance effects for positive, negative number comparisons.

Location of peak voxel	MNI coordinates			Cluster size	Peak t
	x	y	z		
FIFTH GRADERS POSITIVE DISTANCE EFFECT					
L middle, inferior frontal gyrus	−36	35	40	372	5.69
R superior, middle frontal gyrus	24	50	1	615	4.6
L superior, middle frontal gyrus	−27	53	1	85	4.33
R superior frontal gyrus	9	26	61	26	2.75
R middle, inferior frontal gyrus	24	26	−20	23	2.71
R middle frontal gyrus	45	−1	61	18	2.68
L superior parietal lobule	−33	−58	58	11	2.59
L inferior frontal gyrus	−57	14	19	23	2.38
L inferior frontal gyrus	−42	44	1	27	2.38
FIFTH GRADERS NEGATIVE DISTANCE EFFECT					
Bilateral superior, right middle frontal gyrus	0	14	55	304	4.51
L superior frontal gyrus	−18	50	34	123	4.14
R inferior frontal gyrus	36	23	−11	141	3.93
R superior frontal gyrus	24	53	−11	75	3.67
L superior, middle frontal gyrus	−9	29	61	320	3.66
R superior frontal gyrus	6	8	67	54	3.33
R middle frontal gyrus	54	38	22	43	3.25
R inferior parietal lobule	51	−55	43	79	3.15
L inferior frontal gyrus	−48	−1	22	43	3.13
R inferior frontal gyrus	51	2	25	89	2.98
L middle, superior frontal gyrus	−24	59	−11	248	2.96
R inferior frontal gyrus	45	35	7	159	2.9
L superior frontal gyrus	−18	−19	70	73	2.75
R superior frontal gyrus	12	44	49	15	2.64
R middle, superior frontal gyrus	24	50	25	126	2.55
L inferior parietal lobule	−42	−43	22	22	2.45
R superior frontal gyrus	9	−19	73	10	2.39
L inferior frontal gyrus	−54	23	4	19	2.32
L inferior frontal gyrus	−48	29	16	15	2.16
FIFTH GRADERS POSITIVE > NEGATIVE DISTANCE EFFECT					
R inferior parietal lobule	45	−67	37	29	3.1
R inferior frontal gyrus	18	20	−20	21	2.79
R middle frontal gyrus	−36	56	25	11	2.56
R superior frontal gyrus	3	62	1	10	2.34
R superior frontal gyrus	24	65	22	11	2.07
FIFTH GRADERS NEGATIVE > POSITIVE DISTANCE EFFECT					
L middle frontal gyrus	−33	2	49	61	4.36
L superior frontal gyrus	−6	5	55	48	4.07
R inferior, middle frontal gyrus	51	44	−14	153	3.65
L inferior frontal gyrus	−39	20	−14	168	3.37
L superior frontal gyrus	−12	38	52	84	3.2
R middle frontal gyrus	24	−4	49	73	3.17
R superior frontal gyrus	15	17	49	16	3.01
L inferior frontal gyrus	−39	38	10	43	2.77
R inferior frontal gyrus	45	38	4	23	2.54
L superior frontal gyrus	−9	−4	70	13	2.52
R superior frontal gyrus	15	53	31	31	2.43
R superior frontal gyrus	12	−1	70	29	2.4
L inferior parietal lobule	−45	−37	46	24	2.33
R inferior frontal gyrus	60	11	28	10	2.22

FIGURE 6 | Fifth-grader neural distance effects. Both positive **(A)** and negative **(B)** number comparisons showed a typical distance effect in the frontal and parietal lobes, though for negative pairs this was only in the right IPS. This effect was greater for positive comparisons in the parietal lobe, but greater for negative pairs in the frontal lobe **(C)**.
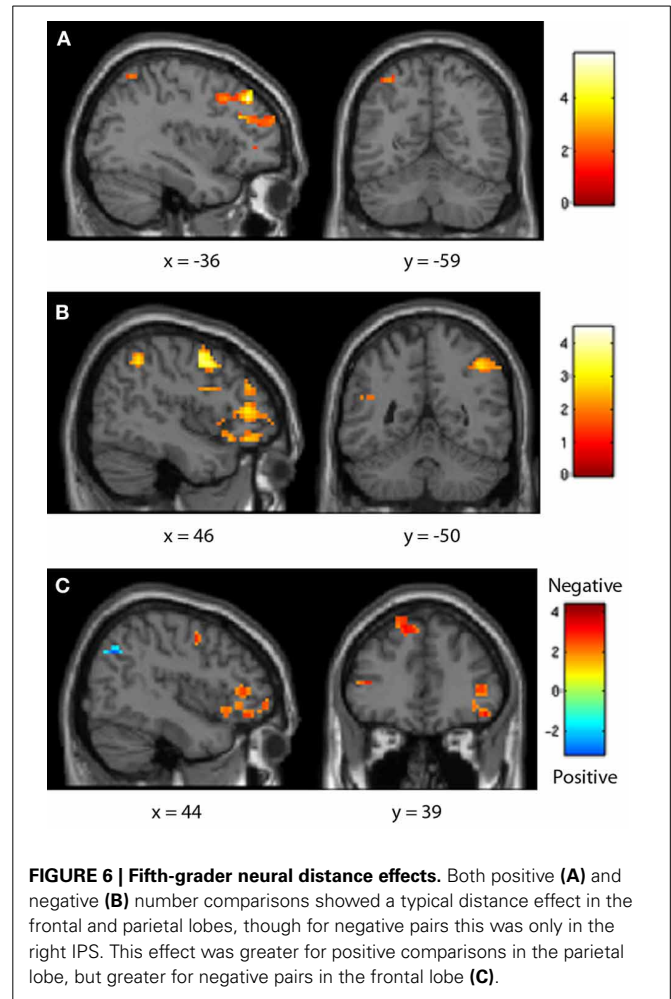
comparisons in the frontal lobe (including the bilateral precentral and inferior frontal gyri) (see **Table 4**, **Figure 6**).

In seventh-graders, positive comparisons demonstrated a typical distance effect both frontally (including the bilateral precentral gyrus and left inferior frontal gyrus) and parietally (including the bilateral IPS). Negative comparisons showed a typical distance effect in the parietal lobe (including the bilateral IPS). This effect was greater for negative comparisons in the left IPS, but was greater for positive pairs in the frontal lobe (including the bilateral inferior frontal gyrus and left precentral gyrus, see **Table 5**, **Figure 7**).

The interaction between age group and the distance effect in each sign was then investigated using a 2 (sign) × 2 (age group) ANOVA (see **Table 6**, **Figure 8**). There were significant effects in the bilateral parietal lobule, including the IPS, as well as the bilateral frontal lobes, especially the inferior and middle frontal gyri (see **Figure 8A**). Differences in the distance effect in each sign between age groups were then investigated by comparing positive greater than negative number distance effects for fifth-greater than seventh-graders. Fifth-graders showed a greater difference between positive and negative number distance effects than seventh-graders in the bilateral IPS and inferior frontal gyrus (see **Figure 8B**).

Table 5 | Seventh-grader neural distance effects for positive, negative comparisons.

Location of peak voxel	MNI coordinates			Cluster size	Peak t
	x	**y**	**z**		
SEVENTH GRADERS POSITIVE DISTANCE EFFECT					
R inferior frontal gyrus	12	38	−20	89	5.92
Bilateral superior frontal gyrus	15	−16	76	59	4.95
L middle, superior frontal gyrus	−39	29	52	122	4.15
L inferior frontal gyrus	−57	38	4	17	3.07
R middle frontal gyrus	45	−4	58	39	2.89
L superior, inferior parietal lobule	−36	−67	43	53	2.84
L inferior parietal lobule	−63	−34	34	32	2.76
R inferior parietal lobule	42	−43	46	27	2.47
L middle frontal gyrus	−42	56	4	28	2.45
L inferior parietal lobule	−45	−40	40	16	2.43
L inferior frontal gyrus	−33	26	−14	10	2.35
L middle frontal gyrus	−36	50	34	14	2.22
SEVENTH GRADERS NEGATIVE DISTANCE EFFECT					
R inferior parietal lobule	45	−49	28	38	3.37
L inferior parietal lobule	−63	−28	31	63	3.22
L inferior parietal lobule	−42	−49	52	60	3.21
R inferior parietal lobule	36	−34	40	56	3.08
R middle frontal gyrus	51	17	43	95	2.71
R middle frontal gyrus	30	−10	49	18	2.71
R superior frontal gyrus	18	17	58	14	2.65
R superior parietal lobule	39	−43	58	31	2.3
R superior parietal lobule	45	−52	52	11	2.22
SEVENTH GRADERS POSITIVE > NEGATIVE DISTANCE EFFECT					
L inferior, superior parietal lobule	−39	−73	43	77	4.84
L inferior frontal gyrus	−33	35	−17	245	4.16
L inferior frontal gyrus	−60	17	31	66	4.09
L middle, superior frontal gyrus	−48	5	55	201	4.01
R superior frontal gyrus	24	65	−8	20	2.94
R superior frontal gyrus	15	−16	79	10	2.78
R inferior frontal gyrus	57	38	−11	11	2.5
L superior frontal gyrus	−12	65	4	10	2.39
R inferior frontal gyrus	54	17	25	13	2.27
SEVENTH GRADERS NEGATIVE > POSITIVE DISTANCE EFFECT					
L inferior parietal lobule	−60	−28	28	17	3.29
R middle frontal gyrus	48	26	43	79	3.18
R inferior frontal gyrus	18	11	−20	11	2.72
R superior frontal gyrus	9	53	37	24	2.57
R inferior parietal lobule	−39	−43	52	18	2.22

FIGURE 7 | Seventh-grader neural distance effects. Both positive **(A)** and negative **(B)** comparisons demonstrated a typical distance effect in the parietal lobe. This effect was greater for positive comparisons in the right IPS and frontal lobe, but greater for negative pairs in the left IPS **(C)**.

COVARIATE EFFECTS ON NEGATIVE NUMBER NEURAL DISTANCE EFFECTS

As comparison accuracy was found to be predictive of the size of the negative number distance effect behaviorally, these relationships were investigated neurally. The impacts of age and task accuracy on the negative number neural distance effect were assessed by including fifth- and seventh-grader negative number distance effect contrasts in the same regression analysis. Participant age (in months) and negative number comparison response accuracy were added as covariates of interest; age and accuracy were not significantly correlated ($r = 0.299$, $p > 0.1$), and so separable effects may be discussed. Across groups, there was a significant neural distance effect for negative number comparisons, with the bilateral IPS demonstrating increasing activity given decreasing comparison distance. As participant age increased, activity increased in the bilateral IPS and decreased in the bilateral inferior parietal lobule and bilateral inferior frontal gyrus and precentral gyrus. As task accuracy increased, activity increased in the left IPS, and decreased in the right IPS and bilateral caudate (see **Table 7**, **Figure 9**). As such, maturation may promote the frontoparietal shift previously noted, but increased task accuracy may either indicate or cause a laterality shift (right to left) within the parietal lobe.
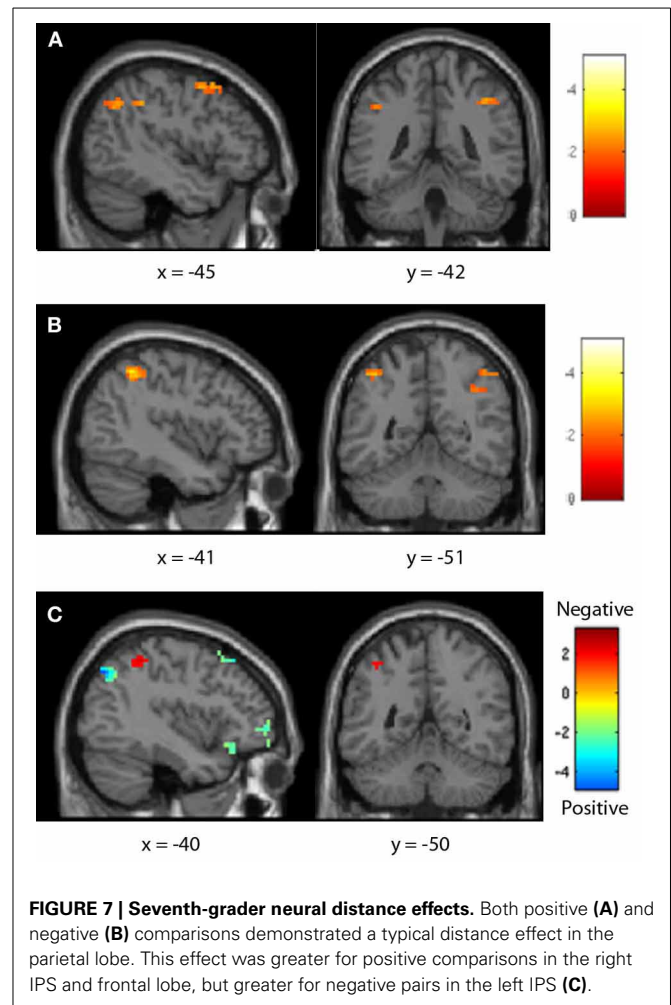
DISCUSSION

The goal of this study was to begin to describe the neural systems involved in negative number processing before, and soon after,

Table 6 | Interaction between sign and age group on the distance effect.

Location of peak voxel	MNI coordinates			Cluster size	Peak t
	x	y	z		
INTERACTION SIGN × AGE GROUP					
R middle frontal gyrus	42	−4	49	128	32.9
R inferior, middle frontal gyrus	60	20	25	321	23.04
R middle, superior frontal gyrus	24	62	−11	19	20.29
L inferior, middle frontal gyrus	−48	14	−2	308	18.94
R inferior frontal gyrus	51	38	−14	45	17.41
R superior frontal gyrus	9	−19	76	76	16.88
R middle frontal gyrus	42	53	1	343	16.71
L superior frontal gyrus	−6	29	49	66	15.26
R superior, inferior parietal lobule	24	−64	46	62	15.13
R middle frontal gyrus	−24	8	49	78	14.35
R inferior frontal gyrus	−33	23	−11	67	14.31
Bilateral superior frontal gyrus	0	11	55	46	13.62
R inferior frontal gyrus	21	17	−23	45	13.4
R middle frontal gyrus	27	35	49	58	12.25
L inferior parietal lobule	−60	−28	34	17	12.16
R superior parietal lobule	33	−49	58	35	11.49
R inferior parietal lobule	39	−37	25	46	9.93
R superior frontal gyrus	12	62	28	14	9.93
R superior, inferior parietal lobule	−30	−70	43	65	9.53
R superior frontal gyrus	21	68	10	33	8.89
R middle frontal gyrus	33	20	37	48	6.13
L middle frontal gyrus	−39	17	40	13	5.57
FIFTH > SEVENTH GRADERS POSITIVE > NEGATIVE NUMBER DISTANCE EFFECT					
L inferior parietal lobule	−63	−34	43	282	4.44
Bilateral superior frontal gyrus	9	5	61	69	4.37
R superior frontal gyrus	24	−7	55	37	4.06
Bilateral superior frontal gyrus	−9	−1	73	129	3.92
R middle frontal gyrus	48	2	43	122	3.86
L middle, inferior frontal gyrus	−60	−1	46	41	3.64
R inferior parietal lobule	45	−49	28	341	3.64
L inferior parietal lobule	−36	−49	25	19	3.59
R inferior frontal gyrus	57	23	−11	14	3.34
R superior parietal lobule	24	−46	64	19	3.22
L superior parietal lobule	−18	−67	55	23	3.19
R superior parietal lobule	27	−64	52	28	3.08
R inferior frontal gyrus	42	56	−5	28	3.01
R inferior frontal gyrus	30	23	−23	20	2.89
L middle frontal gyrus	−33	−1	49	49	2.47
L inferior frontal gyrus	−39	11	−17	10	2.34
L inferior frontal gyrus	−45	17	−8	30	2.21

FIGURE 8 | Age by sign neural distance effect interaction. (A) The bilateral intraparietal sulcus and parts of the inferior and middle frontal gyri were sensitive to the interaction of age group and comparison sign distance effects. **(B)** Fifth-graders showed a greater difference between positive and negative number distance effects in the bilateral IPS and bilateral inferior frontal gyrus than did seventh-graders.

formal instruction on the topic. Pre-instruction fifth-graders and post-instruction seventh-graders were included to examine the effects of age and knowledge level on negative number use. Examinations were primarily conducted between positive and negative number comparison effects to determine whether negative number processing used the same systems, in the same manner, as positives, in each age group.

POSITIVE VERSUS NEGATIVE NUMBER COMPARISONS

Generally, children's behavioral performance was similar to that of the adults previously reported in Gullick et al. (2012): the pattern of positive responses being better (faster, more accurate) than negatives was conserved across ages. Pre-instruction fifth-graders showed reasonable task accuracies and response times for negative numbers, which may indicate some ability to sequence and work with negative numbers even before formal school experience.

The neural contrasts between positive and negative comparisons showed quite different effects across age groups. Seventh-graders showed similar activities for positive and negative comparisons across distances: negative pairs evoked more activity in the left inferior frontal gyrus and bilateral superior temporal gyri, but few other differences were seen, potentially indicating similar representations of the two signs after some instruction. Fifth-graders, though, showed significantly more parietal primary number area activity, including the IPS, for positive than for negative comparisons, and more frontal activity for negatives than positives. This reversed effect demonstrates that the parietal differences seen may not be due simply to difficulty discrepancies between the signs: negative comparisons were harder (slower, less accurate) for fifth-graders, and yet showed decreased parietal activity relative to the easier positive comparisons. Instead, fifth-graders may not yet treat negative numbers as fully "numeric" or quantitatively meaningful, thus limiting their activity in primary quantity-sensitive regions.

Table 7 | Covariate effects on the negative number distance effect across children.

Location of peak voxel	MNI coordinates			Cluster size	Peak t
	x	y	z		
NEGATIVE DISTANCE EFFECT ACROSS PARTICIPANTS					
L middle frontal gyrus	−36	8	52	117	3.85
R middle frontal gyrus	27	−7	52	357	3.59
R middle, superior frontal gyrus	24	50	−11	90	3.34
R inferior parietal lobule	57	−52	46	182	3.2
R superior frontal gyrus	12	2	70	94	3.03
R inferior parietal lobule	48	−46	25	52	2.96
L inferior parietal lobule	−39	−46	49	82	2.79
R superior frontal gyrus	18	14	49	59	2.75
R inferior frontal gyrus	33	23	−11	163	2.75
L inferior parietal lobule	−42	−46	22	33	2.7
L inferior frontal gyrus	−48	−1	22	29	2.58
L superior frontal gyrus	−9	−19	70	23	2.57
L superior frontal gyrus	−6	−4	70	33	2.57
L superior frontal gyrus	−57	11	43	10	2.56
R superior frontal gyrus	15	53	34	42	2.54
R inferior parietal lobule	45	−25	28	10	2.36
L superior, middle frontal gyrus	−15	44	34	59	2.29
R inferior frontal gyrus	48	−1	25	14	2.25
L inferior frontal gyrus	−48	14	−2	24	2.1
INCREASING ACTIVITY WITH INCREASING AGE					
L superior parietal lobule	−12	−67	58	28	2.85
R middle frontal gyrus	36	8	37	24	2.72
R inferior parietal lobule	39	−40	34	23	2.72
L inferior, superior parietal lobule	−39	−52	55	46	2.53
R inferior parietal lobule	30	−43	58	35	2.28
R superior frontal gyrus	18	14	58	14	2.23
DECREASING ACTIVITY WITH INCREASING AGE					
L superior, middle frontal gyrus	−12	59	−11	273	3.97
R middle, superior frontal gyrus	21	65	−11	53	3.7
L middle, superior frontal gyrus	−27	8	49	273	3.37
R inferior frontal gyrus	51	41	−14	35	3.04
R inferior frontal gyrus	57	17	25	61	2.78
R superior, middle frontal gyrus	15	17	49	115	2.77
R middle frontal gyrus	24	−1	49	31	2.69
L superior frontal gyrus	−21	59	19	54	2.68
L inferior parietal lobule	−48	−37	22	25	2.66
L inferior parietal lobule	−54	−58	40	32	2.64
L middle frontal gyrus	−39	−1	61	16	2.44
L inferior frontal gyrus	−54	17	4	20	2.4
R inferior frontal gyrus	33	11	−20	12	2.36
L middle, superior frontal gyrus	−21	50	34	19	2.35
R middle frontal gyrus	48	56	−2	12	2.25
R inferior frontal gyrus	39	38	7	10	2.02
INCREASING ACTIVITY WITH INCREASING NEGATIVE NUMBER ACCURACY					
R superior frontal gyrus	18	65	−11	23	3.5
R inferior parietal lobule	54	−64	40	49	2.7
L middle frontal gyrus	−42	56	10	12	2.46

(Continued)

Table 7 | Continued

Location of peak voxel	MNI coordinates			Cluster size	Peak t
	x	y	z		
L middle frontal gyrus	−30	62	−14	18	2.44
R middle frontal gyrus	45	56	10	12	2.38
L inferior parietal lobule	−42	−67	43	72	2.34
DECREASING ACTIVITY WITH INCREASING NEGATIVE NUMBER ACCURACY					
L superior frontal gyrus	−18	50	4	61	4.04
R middle, superior frontal gyrus	51	38	−2	479	3.99
R middle, superior frontal gyrus	36	35	22	237	3.62
L inferior frontal gyrus	−57	41	7	19	3.07
R inferior parietal lobule	42	−40	34	55	3.02
R superior, middle frontal gyrus	21	32	46	108	2.99
L inferior, middle frontal gyrus	−24	29	−5	142	2.89
L middle frontal gyrus	−42	23	52	15	2.87
L superior frontal gyrus	−15	59	31	63	2.64
L middle frontal gyrus	−36	14	43	25	2.51
R inferior parietal lobule	39	−49	52	26	2.45
L inferior frontal gyrus	−30	11	−14	23	2.42
R inferior frontal gyrus	63	11	25	30	2.28
R superior frontal gyrus	24	−13	64	20	2.22
L superior frontal gyrus	−12	41	37	10	2.13

FIGURE 9 | Negative number neural distance effects across children. Fifth- and seventh-graders together demonstrated a bilateral IPS distance effect for negative number comparisons. **(A)** A frontoparietal shift was seen with increasing age **(B)**, and a right-to-left laterality shift was seen in the IPS with increasing negative comparison accuracy **(C)**.
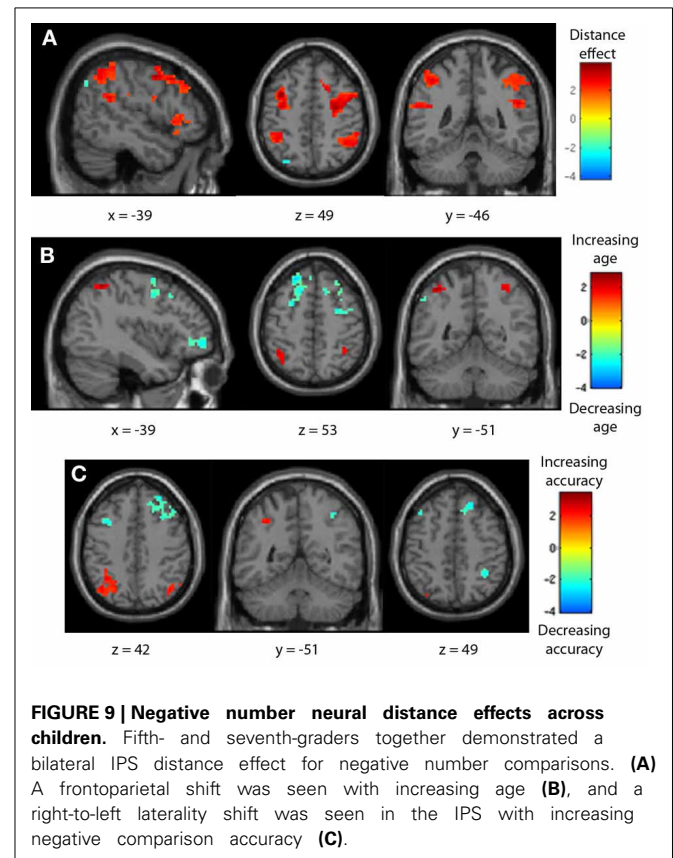
POSITIVE AND NEGATIVE NUMBER DISTANCE EFFECTS

Both age groups demonstrated typical-direction behavioral distance effects for negative number comparisons. Developmentally, the size of the distance effect decreases with age, and a smaller distance effect may indicate a more mature representation of number (Holloway and Ansari, 2008). The negative number distance effect in adults has been previously demonstrated to be larger than that for positive, demonstrating that even after years of practice the left side of the mental number line is not quite as mature as the right (Gullick et al., 2012). The increased effect seen in children may also support this idea, as these participants have received only limited instruction or informal practice with negatives.

Behaviorally, positive number comparisons did not demonstrate a significant distance effect in the context-collapsed analyses; however, context-specific analyses demonstrated a typical distance effect for digit-format comparisons. While thermometer-format presentations appeared to aid processing of negative numbers, especially for the relatively naïve fifth graders, they seemed to interfere with positive number processing for both age groups. Unfortunately, we did not have enough statistical power to compare fMRI data between presentation contexts, and instead collapsed across digit- and thermometer-format trials. As such, there may be interesting neural differences between thermometer and digit comparison processing, especially given the context by age effects on negative number behavioral distance effects, but we cannot investigate these questions here. Further work is needed to better understand the impact of presentation context on number processing for both already-known positive numbers and recently-introduced negative numbers.

Neurally, all age groups demonstrated a typical distance effect for positive comparisons in the IPS, as well as in the precentral and inferior frontal gyrus. Previously, Ansari et al. (2005) noted only a right-lateralized frontal distance effect for children ages 8–12; while parietal regions were sensitive to number, their activity did not significantly differ between close and far comparison distances. Distance effects were proposed to shift at some point from the child-like frontal regions to the adult-like parietal regions, but the timeline of this shift was not more specifically defined. The children included in the current study fall in the upper end of this age range, and did show a parietal distance effect, as well as a right-lateralized precentral gyrus effect similar to that noted by Ansari et al. (2005) and a left-lateralized inferior frontal gyrus effect. Further, Ansari et al. (2005) categorized comparison distance as "close" or "far" and contrasted the two, whereas the present study used distance as a continuous parametric regressor, which may be more sensitive to small but important changes in processing. This study thus helps to better track the developmental trajectory of the frontoparietal shift even for positive number usage.

Similarly, both age groups demonstrated a typical distance effect for negative comparisons in the IPS and the precentral and inferior frontal gyri. Negative numbers thus draw on quantity-sensitive regions in a manner similar to that for positive

comparisons, even in pre-instruction children. However, positive number comparisons demonstrated a greater distance effect in the IPS than negative comparisons in pre-instruction children; post-instruction children showed fewer neural differences between positive and negative number distance effects. As such, fifth-graders showed a greater difference between distance effects in each sign than did seventh-graders. These results may again indicate that while negative numbers may use primary number regions, it is not to the same degree as positives, at least before formal instruction and practice have occurred. This difference may again imply a less mature representation of negative numbers' quantities before instruction.

INDIVIDUAL VARIABILITY IN THE NEGATIVE NUMBER DISTANCE EFFECT

While on average fifth-grade children demonstrated a particularly immature distance effect for negative numbers, there was a large amount of individual variability within the group: some children did not show any consistent effect of distance on negative comparison response time, and some showed a large effect. Though a larger distance effect may indicate relative immaturity of or decreased precision in numeric representations, no distance effect for negative number comparisons at all may more indicate that negatives are not ordered or sequentially arranged. Fifth-graders' negative number distance effect was predicted only by their accuracy on positive number comparisons: children with low positive comparison accuracy tended to show a very small or reversed distance effect for negatives (implying a lack of organization), but higher-accuracy children were more likely to show a typical-direction distance effect. More broadly, children who were better at working with numbers generally were better at working with negative numbers specifically. A better understanding of quantity overall may provide a stronger base from which to work with difficult concepts like negative numbers, leading to a more mature effect. This finding suggests that, before instruction on this difficult concept, use of negatives may rely on one's ability to work with positives.

In contrast, seventh-graders' negative number comparison distance effect was predicted by negative comparison accuracy and Stroop interference score. First, children with higher negative comparison accuracies were more likely to show a typical distance effect for negative numbers. After instruction on negative numbers, then, the ability to use negatives may be based more on one's understanding of negatives themselves, and not on positive numbers. This difference between ages may represent a shift in understanding, though the mechanism behind this change cannot yet be discerned.

Second, the impact of the Stroop interference score on seventh-grader negative number distance effects may indicate the inhibition necessary in responding to negative numbers: to choose the greater number, participants must pick the smaller digit, which may be especially difficult in closer comparisons with a smaller difference between the values. However, the restriction of this effect to only the seventh-graders may indicate the prerequisite of some knowledge of negatives for inhibition to differentially effect responses across comparison distances.

As this study is cross-sectional, and occurred in a specific area where negative numbers are taught in the same grade across schools, it necessarily confounds participant age with instruction. Only the older seventh-grade children had learned about negative numbers in school, while none of the younger fifth-graders had. Luckily, the experimental sample included a range of ages and abilities within each grade group, making it possible to separately examine the impact of these two factors on brain activity and negative number performance. Both groups together demonstrated an expected parietal (IPS) and frontal (inferior frontal gyrus, precentral gyrus) distance effect for negative number comparisons. Changes in participant age were associated with the frontoparietal shift previously noted: negative number frontal distance effects decreased and parietal distance effects increased as participant age increased. This finding may again reflect the general developmental trend of a frontoparietal shift in numerical cognition (Ansari et al., 2005; Rivera et al., 2005). Interestingly, Rosenberg-Lee et al. (2011) reported a non-linear increase in both frontal and parietal activity in children from second to third grade in arithmetic problem solving, perhaps similar to the non-linear shifts seen here.

Alternately, the increased frontal distance effect for fifth-graders may be the result of relatively heightened strategy implementation. Pre-instruction children may primarily understand negatives through the use of rules, such as sign changes, which transform unfamiliar negative numbers into known positive values (see Varma and Schwartz, 2011); these rules may be especially important in solving close-distance pairs. With further instruction and experience, though, these rules may become less necessary, giving rise to the age-related differences seen.

Changes in participant accuracy on negative number comparisons, though, were associated with a right- to left-hemisphere IPS change. As participants' ability to work with negative numbers and to treat them as quantitative values increases, the distance effect for negatives seems to increase in the left IPS and decrease in the right. This pattern is also seen between age groups in the transition from a right IPS only negative number distance effect in fifth-graders to the bilateral effect seen in seventh-graders. This laterality shift may align with previously proposed separations within the mental number system. The right IPS has been noted to be more responsive to non-symbolic than symbolic number representations, and has been hypothesized to particularly support approximate quantitative processing. The left IPS, in contrast, is responsive to quantitative information across notations but is especially sensitive to practiced, encultured symbolic numbers (Ansari, 2007; Kadosh et al., 2007; Piazza et al.,

2007) and thus may be more able to represent exact amounts via symbolic cognition.

Within this comparison task, then, low performers may be more likely to understand negative numbers approximately, representing them as approximately small values via the right IPS, and thus also demonstrating larger behavioral distance effects. High-performers (and adults) may be more able to represent negative numbers as precise values, thus demonstrating increased distance-related activity in the left IPS and more mature distance effects. Whether the ability to represent negative values precisely causes or follows the shift cannot be determined from this study, but it does at least appear to be related.

SUMMARY

This study thus presents a first examination of the neural correlates of negative number processing in pre- and post-instruction children. As younger participants were expected to demonstrate a more frontal-based distance effect, and older possibly a more parietal-focused effect, neural analyses were first conducted separately within each age group to better directly compare the activities found in positive and negative pairs. Pre-instruction fifth-grader responses were, on average, reasonably accurate, demonstrating some knowledge of negative numbers even before formal instruction, and showed similar behavioral response patterns to older participants, indicating typical number line arrangement and use. All participants demonstrated a significant neural distance effect for both positive and negative number comparisons in the IPS, though frontal effects were also seen. Even before instruction, then, children may be able to treat negative numbers as representations of quantitative values and draw on number-related brain regions, even if in a relatively immature manner. Increasing age demonstrated a significant frontoparietal shift consistent with previous developmental numerical cognition work, but increasing negative comparison accuracy showed a right IPS to left IPS shift, possibly indicating a change from approximate to precise negative quantity representations. These shifts and changes illustrate the process of incorporation of negative numbers as quantitative entities into the mental number system across years of practice and experience.

REFERENCES

Anderson, J. R., Betts, S., Ferris, J. L., and Fincham, J. M. (2011). Cognitive and metacognitive activity in mathematical problem solving: prefrontal and parietal patterns. *Cogn. Affect. Behav. Neurosci.* 11, 52–67. doi: 10.3758/S13415-010-0011-0

Ansari, D. (2007). Does the parietal cortex distinguish between "10," "ten,"

and ten dots? *Neuron* 53, 165–167. doi: 10.1016/j.neuron.2007.01.001

Ansari, D., and Dhital, B. (2006). Age-related changes in the activation of the intraparietal sulcus during nonsymbolic magnitude processing: an event-related functional magnetic resonance imaging study. *J. Cogn. Neurosci.* 18, 1820–1828. doi: 10.1162/jocn.2006.18.11.1820

Ansari, D., Dhital, B., and Siong, S. C. (2006). Parametric effects of numerical distance on the intraparietal sulcus during passive viewing of rapid numerosity changes. *Brain Res.* 1067, 181–188. doi: 10.1016/j.brainres.2005.10.083

Ansari, D., Garcia, N., Lucas, E., Hamon, K., and Dhital, B. (2005). Neural correlates of symbolic

number processing in children and adults. *Neuroreport* 16, 1769–1773.

Barth, H., Beckmann, L., and Spelke, E. S. (2008). Nonsymbolic, approximate arithmetic in children: abstract addition prior to instruction. *Dev. Psychol.* 44, 1466–1477. doi: 10.1037/a0013046

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N.,

and Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition* 98, 199–222. doi: 10.1016/j.cognition.2004.09.011

Barth, H., La Mont, K., Lipton, J., and Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14116–14121. doi: 10.1073/pnas.0505512102

Blair, K., Rosenberg-Lee, M., Tsang, J. M., Schwartz, D., and Menon, V. (2012). Beyond natural numbers: representation of negative numbers in the parietal cortex. *Front. Hum. Neurosci.* 6:7. doi: 10.3389/fnhum.2012.00007

Borba, R., and Nunes, T. (1999). Young children's representations of negative numbers. *Proc. Br. Soc. Res. Learn. Math.* 19, 7–12.

Brett, M. (2006). *MniTalairach - MRC CBU Imaging Wiki*. Available online at: http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach

Burgund, E. D., Kang, H. C., Kelly, J. E., Buckner, R. L., Snyder, A. Z., and Petersen, S. E. (2002). The feasibility of a common stereotactic space for children and adults in fMRI studies of development. *Neuroimage* 17, 184–200. doi: 10.1006/nimg.2002.1174

Cantlon, J. F., Brannon, E. M., Carter, E. J., and Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-old children. *PLoS Biol.* 4:844–854. doi: 10.1371/journal.pbio.0040125

Cantlon, J. F., Libertus, M. E., Pinel, P., Dehaene, S., Brannon, E. M., and Pelphrey, K. A. (2009). The neural development of an abstract concept of number. *J. Cogn. Neurosci.* 21, 2217–2229. doi: 10.1162/jocn.2008.21159

Caviness, V. S., Kennedy, D. N., Richelme, C., Rademacher, J., and Filipek, P. A. (1996). The human brain age 7-11 years: a volumetric analysis based on magnetic resonance images. *Cereb. Cortex* 6, 726–736. doi: 10.1093/cercor/6.5.726

Chassy, P., and Grodd, W. (2012). Comparison of quantities: core and format-dependent regions as revealed by fMRI. *Cereb. Cortex* 22, 1420–1430. doi: 10.1093/cercor/Bhr219

Cui, X., Li, J., and Song, X. (2011). *xjview: a viewing program for SPM*. Available online at: http://www.alivelearn.net/xjview8/

Davis, R. B., and Maher, C. A. (1993). "The reality of negative numbers," in *Schools, Mathematics, and The World of Reality*, eds R. B. Davis and

C. A. Maher (Boston, MA: Allyn and Bacon), 51–60.

De Smedt, B., Verschaffel, L., and Ghesquiere, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *J. Exp. Child Psychol.* 103, 469–479. doi: 10.1016/j.jecp.2009.01.010

Education, N. H. D. O. (2010). *Common Core State Standards for Mathematics*, Concord, NH.

Fischer, M. H. (2003). Cognitive representation of negative numbers. *Psychol. Sci.* 14, 278–282. doi: 10.1111/1467-9280.03435

Ganor-Stern, D., Pinhas, M., Kallai, A., and Tzelgov, J. (2010). Holistic representation of negative numbers is formed when needed for the task. *Q. J. Exp. Psychol.* 63, 1969–1981. doi: 10.1080/17470211003721667

Ganor-Stern, D., and Tzelgov, J. (2008). Negative numbers are generated in the mind. *Exp. Psychol.* 55, 157–163. doi: 10.1027/1618-3169.55.3.157

Gilmore, C. K., McCarthy, S. E., and Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature* 447, 589–591. doi: 10.1038/nature05850

Gilmore, C. K., and Spelke, E. S. (2008). Children's understanding of the relationship between addition and subtraction. *Cognition* 107, 932–945. doi: 10.1016/j.cognition.2007.12.007

Gobel, S. M., Johansen-Berg, H., Behrens, T., and Rushworth, M. F. S. (2004). Response-selection-related parietal activation during number comparison. *J. Cogn. Neurosci.* 16, 1536–1551. doi: 10.1162/0898929042568442

Golden, C. J., Freshwater, S. M., and Golden, Z. (2003). *Stroop Color and Word Test: Children's Version for Ages*. Wood Dale, IL: Stoelting, 5–14.

Gullick, M. M., and Wolford, G. (2013). Brain systems involved in arithmetic with positive versus negative numbers. *Hum. Brain Mapp.* doi: 10.1002/hbm.22201

Gullick, M. M., Wolford, G., and Temple, E. (2012). Understanding less than nothing: neural distance effects for negative numbers. *Neuroimage* 62, 542–554. doi: 10.1016/j.neuroimage.2012.04.058

Holloway, I. D., and Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. *Dev. Sci.* 11, 644–649. doi: 10.1111/j.1467-7687.2008.00712.X

Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes

onto symbols: the numerical distance effect and individual differences in children's mathematics achievement. *J. Exp. Child Psychol.* 103, 17–29. doi: 10.1016/j.jecp.2008.04.001

Holloway, I. D., and Ansari, D. (2010). Developmental specialization in the right intraparietal sulcus for the abstact representation of numerical magnitude. *J. Cogn. Neurosci.* 22, 2627–2637. doi: 10.1162/jocn.2009.21399

Holloway, I. D., Price, G. R., and Ansari, D. (2010). Common and segregated neural pathways for the processing of symbolic and nonsymbolic numerical magnitude: an fMRI study. *Neuroimage* 49, 1006–1017. doi: 10.1016/j.neuroimage.2009.07.071

Kadosh, R. C., Kadosh, K. C., Kaas, A., Henik, A., and Goebel, R. (2007). Notation-dependent and -independent representations of numbers in the parietal lobes. *Neuron* 53, 307–314. doi: 10.1016/j.neuron.2006.12.025

Kang, H. C., Burgund, E. D., Lugar, H. M., Petersen, S. E., and Schlaggar, B. L. (2003). Comparison of functional activation foci in children and adults using a common stereotactic space. *Neuroimage* 19, 16–28. doi: 10.1016/S1053-8119(03)00038-7

Kaufman, A. S., and Kaufman, N. L. (2004). *Kaufman Test of Educational Achievement, 2nd Edn. (KTEA-II)*, (San Antonio, TX: Pearson).

Lancaster, J. L., Rainey, L. H., Summerlin, J. L., Freitas, C. S., Fox, P. T., and Evans, A. C. (1997). Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. *Hum. Brain Mapp.* 5, 238–242. doi: 10.1002/(SICI)1097-0193(1997)5:4<238::AID-HBM6>3.0.CO;2-4

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, E. S., and Rainey, L. (2000). Automated Talairach Atlas labels for functional brain mapping. *Hum. Brain Mapp.* 10, 120–131. doi: 10.1002/1097-0193(200007)10:3<120:AID-HBM30>3.0.CO;2-8

Lipton, J. S., and Spelke, E. S. (2003). Origins of number sense: large-number discrimination in human infants. *Psychol. Sci.* 14, 396–401. doi: 10.1111/1467-9280.01453

Lyons, I. M., and Ansari, D. (2009). The cerebral basis of mapping nonsymbolic numerical quantities onto abstract symbols: an fMRI training study. *J. Cogn.*

Neurosci. 21, 1720–1735. doi: 10.1162/jocn.2009.21124

McCrink, K., and Wynn, K. (2007). Ratio abstraction by 6-month-old infants. *Psychol. Sci.* 18, 740–745. doi: 10.1111/j.1467-9280.2007.01969.x

Menon, V., Rivera, S. M., White, C. D., Eliez, S., Glover, G. H., and Reiss, A. L. (2000). Functional optimization of arithmetic processing in perfect performers. *Cogn. Brain Res.* 9, 343–345. doi: 10.1016/S0926-6410(00)00010-0

Menon, V., Rivera, S. M., White, C. D., Glover, G. H., and Reiss, A. L. (2000). Dissociating prefrontal and parietal cortex activation during arithmetic processing. *Neuroimage* 12, 357–365. doi: 10.1006/nimg.2000.0613

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

Mukhopadhyay, S., Resnick, L. B., and Schauble, L. (1990). "Social sense-making in mathematics: children's ideas of negative numbers," in *Paper Presented at the Psychology of Mathematics Education Conference*, (Mexico).

Oldfield, R. C. (1971). Assessment and analysis of handedness - edinburgh inventory. *Neuropsychologia* 9, 97.

Piazza, M., Pinel, P., Le Bihan, D., and Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron* 53, 293–305. doi: 10.1016/j.neuron.2006.11.022

Pinel, P., Dehaene, S., Riviere, D., and LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *Neuroimage* 14, 1013–1026. doi: 10.1006/nimg.2001.0913

Poldrack, R. A., Mumford, J. A., and Nichols, T. E. (2011). *Handbook of Functional MRI Data Analysis*. New York, NY: Cambridge University Press.

Poline, J. B., Kherif, F., and Penny, W. (2003). "Contrasts and classical inference," in *Human Brain Function, 2nd Edn.*, eds R. Frackowiak, K. J. Friston, C. Frith, R. Dolan, C. J. Price, S. Zeki, J. Ashburner and W. Penny (London: Academic Press), 761–780.

Rivera, S. M., Reiss, A. L., Eckert, M. A., and Menon, V. (2005). Developmental changes in mental arithmetic: evidence for increased functional specialization in the left inferior parietal cortex. *Cereb. Cortex* 15, 1779–1790. doi: 10.1093/cercor/bhi055

Rosenberg-Lee, M., Barth, M., and Menon, V. (2011). What difference does a year of schooling make? maturation of brain response and connectivity between 2nd and 3rd grades during arithmetic problem solving. *Neuroimage* 57, 796–808. doi: 10.1016/j.neuroimage.2011.05.013

Roux, F. E., Lubrano, V., Lauwers-Cances, V., Giussani, C., and Demonet, J. F. (2008). Cortical areas involved in Arabic number reading. *Neurology* 70, 210–217. doi: 10.1212/01.wnl.0000297194.14452.a0

Schneider, M., Grabner, R. H., and Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: their interrelations in grades 5 and 6. *J. Educ. Psychol.* 101, 359–372. doi: 10.1037/a0013840

Streefland, L. (1996). Negative numbers: reflections of a learning researcher. *J. Math. Behav.* 15, 57–77.

Temple, E., and Posner, M. I. (1998). Brain mechanisms of quantity are similar in 5-year-old children and adults. *Proc. Natl. Acad. Sci. U.S.A.* 95, 7836–7841.

Tsang, J. M., and Schwartz, D. L. (2009). "Symmetry in the semantic representation of integers," in *Paper presented at the Annual Meeting of the Cognitive Science Society* (Amsterdam).

Tzelgov, J., Ganor-Stern, D., and Maymon-Schreiber, K. (2009). The representation of negative numbers: exploring the effects of mode of processing and notation. *Q. J. Exp. Psychol.* 62, 605–624. doi: 10.1080/17470210802034751

Van Opstal, F., Gevers, W., De Moor, W., and Verguts, T. (2008). Dissecting the symbolic distance effect: comparison and priming effects in numerical and nonnumerical orders. *Psychon. Bull. Rev.* 15, 419–425. doi: 10.3758/pbr.15.2.419

Varma, S., and Schwartz, D. L. (2011). The mental representation of integers: an abstract-to-concrete shift in the understanding of mathematical concepts. *Cognition* 121, 363–385. doi: 10.1016/j.cognition.2011.08.005

Varma, S., Schwartz, D. L., Lindgren, R., and Go, J. (2007). "Integer comparison and the inverse symbolic distance effect," in *Paper presented at the Proceedings of the 19th Annual Cognitive Science Society* (Nashville, TN).

Vlassis, J. (2004). Making sense of the minus sign or becoming flexible in 'negativity.' *Learn. Instruct.* 14, 469–484. doi: 10.1016/j.learninstruc.2004.06.012

Widjaja, W., Stacey, K., and Steinle, V. (2011). Locating negative decimals on the number line: insights into the thinking of pre-service primary teachers. *J. Math. Behav.* 30, 80–91. doi: 10.1016/j.jmathb.2010.11.004

Wynn, K. (1992). Addition and subtraction by human infants. *Nature* 358, 749–750. doi: 10.1038/358749a0

Xu, F., and Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition* 74, B1–B11. doi: 10.1016/S0010-0277(99)00066-9

Xu, F., Spelke, E. S., and Goddard, S. (2005). Number sense in human infants. *Dev. Sci.* 8, 88–101. doi: 10.1111/j.1467-7687.2005.00395.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

Integer Knowledge Test.

Write these numbers with the opposite sign:

1. −7 _____   2. 11 _____   3. 137 _____

4. 4 _____   5. −22 _____   6. −13 _____

Place these numbers on a number line:

1. 4

3. −2

2. 1

4. −5
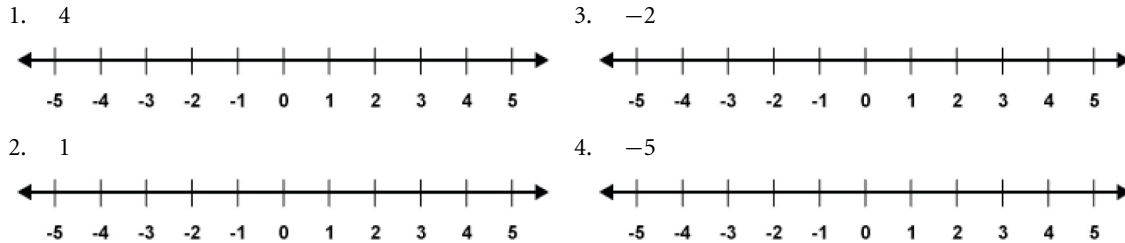
Order these numbers from least to greatest:

1. 0, −4, 4 _____   2. −2, −5, 6, 1 _____

3. −3, 24, 17, −6, 2, 1 _____

4. 19, 17, −17, −19, 0 _____

Circle the greater number:

1. 5 or 9   2. −3 or −8   3. −19 or −12

4. −2 or 6   5. 17 or −14   6. 16 or −21

Add the following integers:

1. 13 + 1 = _____   2. −4 + −6 = _____

3. −2 + −12 = _____   4. −6 + 8 = _____

5. −11 + 8 = _____   6. 2 + −11 = _____

Subtract the following integers:

1. 7 − 3 = _____   2. −10 − 4 = _____

3. 5 − 2 = _____   4. −12 − −5 = _____

5. −6 − 3 = _____   6. 4 − −9 = _____

Multiply these integers:

1. 6 × 7 = _____   2. −4 × −9 = _____

3. −8 × −3 = _____   4. −10 × 5 = _____

5. 6 × −9 = _____   6. −2 × 7 = _____

Divide these integers:

1. 36 ÷ 2 = _____   2. −12 ÷ −4 = _____

3. −45 ÷ −5 = _____   4. −18 ÷ 6 = _____

5. 24 ÷ −3 = _____   6. −90 ÷ 10 = _____

Solve for x in each case:

1. $x - 3 = 12$ _____
2. $48 - x = 31$ _____
3. $x \div -11 = 2$ _____
4. $x/-5 = -8$ _____
5. $x + 8 > 18$ _____
6. $x - 1 < 3$ _____
7. $x + 8 > -4$ _____
8. $x \div 4 < 3$ _____
9. $-5x > 25$ _____
10. $-6x < -42$ _____
11. $-5 < x/4$ _____
12. $x/4 < -1/4$ _____

Write an equation describing each of these word problems, then solve it:

1. Maggie owes the candy store $35. Five of her friends will help her pay off her debt. How much will each friend pay? (Maggie has no money to help pay).

2. The temperature at noon on a winter day was 8°C. At midnight, the temperature had dropped by 15°. What was the temperature at midnight?

3. A drawing of a building shows the elevation of the basement floor to be 12 feet below ground level (–12 feet). The elevation of the roof is 32 feet. What is the total distance from the roof to the basement?

Age does *not* count: resilience of quantity processing in healthy ageing

Anna Lambrechts[1], Vyacheslav Karolis[2], Sara Garcia[3], Jennifer Obende[4] and Marinella Cappelletti[2]*

[1] *Autism Research Group, Department of Psychology, City University London, London, UK*
[2] *Institute of Cognitive Neuroscience, University College London, London, UK*
[3] *Institute of Ophthalmology, University College London, London, UK*
[4] *Psychology Department, University College London, London, UK*

Quantity skills have been extensively studied in terms of their development and pathological decline. Recently, numerosity discrimination (i.e., how many items are in a set) has been shown to be resilient to healthy ageing despite relying on inhibitory skills, but whether processing continuous quantities such as time and space is equally well-maintained in ageing participants is not known. Life-long exposure to quantity-related problems may progressively refine proficiency in quantity tasks, or alternatively quantity skills may decline with age. In addition, is not known whether the tight relationship between quantity dimensions typically shown in their interactions is preserved in ageing. To address these questions, two experimental paradigms were used in 38 younger and 32 older healthy adults who showed typical age-related decline in attention, executive function and memory tasks. In both groups we first assessed time and space discrimination independently using a two-choice task (i.e., "Which of two horizontal lines is longer in duration or extension?"), and found that time and space processing were equally accurate in younger and older participants. In a second paradigm, we assessed the relation between different quantity dimensions which were presented as a dynamic pattern of dots independently changing in duration, spatial extension and numerosity. Younger and older participants again showed a similar profile of interaction between number, cumulative area and duration, although older adults showed a greater sensitivity to task-irrelevant information than younger adults in the cumulative area task but lower sensitivity in the duration task. Continuous quantity processing seems therefore resilient to ageing similar to numerosity and to other non-quantity skills like vocabulary or implicit memory; however, ageing might differentially affect different quantity dimensions.

Keywords: quantity processing, time, space, number, ageing, magnitude system

INTRODUCTION

A central part of our everyday life involves judging quantities, for example which queue at the supermarket has fewer people, or if a parking space is wide enough for our car, or if there is sufficient time to pop to a café before our next meeting (Lemaire and Lecacheur, 2007; Gandini et al., 2008, 2009). Collectively these judgments provide rough magnitude estimates in the form of number, spatial extension or temporal duration (Walsh, 2003; Gandini et al., 2009; Bueti and Walsh, 2009; Bonn and Cantlon, 2012; Cantlon, 2012).

A large body of research has recently investigated the development of numerical, spatial and temporal estimations in humans and primates and their impairment in the lesioned brain. For instance, these studies have shown that different magnitude dimensions have parallel patterns of performance in animals (Meck and Church, 1983; Breukelaar and Dalrymple-Alford, 1998; Meck, 2005; Beran, 2007; Merritt et al., 2010), and similar rates of development in humans (Brannon et al., 2006, 2007; Van Marle and Wynn, 2006; Feigenson, 2007; Droit-volet et al., 2008; Reynvoet et al., 2009). It has also been shown that there are

associations (Basso et al., 1996; Zorzi et al., 2002) and dissociations between dimensions in the lesioned brain (Doricchi et al., 2005; Cappelletti et al., 2009, 2011). This evidence has supported the idea that magnitude dimensions are mapped onto an abstract analogue scale (Walsh, 2003; Bueti and Walsh, 2009; Gallistel, 2011) such that from early in development individuals apply associative mappings "more A, more B" across different magnitude dimensions (Lourenco and Longo, 2010).

MAGNITUDE PROCESSING IN HEALTHY AGEING

Some research has investigated math and numerosity processing in older age, but processing of continuous quantities such as space and time in ageing is less known. Existing studies on number have focused on elderly's mathematical abilities (Halberda et al., 2012; Duverne and Lemaire, 2005; Dormal et al., 2012), and recently on more foundational skills such as numerosity discrimination (Halberda et al., 2012; Cappelletti et al., in press). These studies concurred to show that although older participants can learn new ways to solve arithmetical problems, they show a smaller repertoire of strategies and are less efficient than

younger participants in selecting among them (e.g., Duverne and Lemaire, 2005; Lemaire and Arnaud, 2008), or that they do not equally engage the same brain regions as younger participants when performing arithmetical tasks (El Yagoubi et al., 2005). Moreover, Cappelletti et al. (in press) found that numerosity discrimination is resilient to ageing although it is influenced by the decline of inhibitory processes supporting number performance. In comparison to number, time and space processing have been much less investigated in older adults (OAs). Some evidence in the temporal domain indicates that OAs demonstrate diminished accuracy but intact sensitivity in duration judgments (Baudouin et al., 2006; Block et al., 1998; Lustig and Meck, 2011). For instance, OAs report larger estimates but reproduce shorter durations relative to younger adults (YAs) (Block et al., 1998). However, these group differences might reflect age-related declining of skills required for temporal judgments, like working memory storage and executive functions, in which case they would not be suggestive of a pure deficit in temporal processing in ageing. In the domain of spatial processing, differences in speed (Birren and Botwinick, 1955) but not in accuracy (Verrillo, 1981; but see Sara and Faubert, 2000) have been reported between YAs and OAs in size discrimination tasks.

RELATIONS BETWEEN MAGNITUDE DIMENSIONS FROM INFANCY TO OLDER AGE

One way to probe the integrity of magnitude processing lays in examining interaction effects between magnitude dimensions, and interactions are at the basis of the idea that different dimensions are mapped on an analogue scale, and that magnitude processing skills rely on common resources (Gallistel, 1989, 2011; Gallistel and Gelman, 2000; Walsh, 2003; Bueti and Walsh, 2009; Cantlon, 2012). Interaction studies have shown that judgments on a target dimension are sensitive to information from concurrent task-irrelevant magnitude dimensions. Such studies usually examine the influence of one dimension on the other, unilaterally (effect of A on B) or bilaterally (effect of A on B and effect of B on A). For instance, duration has been recurrently found to be sensitive to task-irrelevant numerical information (both symbolic, like Arabic figures and non-symbolic, like number of dots) following a "more A, more B" pattern: the larger the number, the longer the duration is perceived (Droit-Volet et al., 2003; Dormal et al., 2006; Xuan et al., 2007, 2009; Oliveri et al., 2008; Vicario et al., 2008; Dormal and Pesenti, 2013). Duration processing has also been shown to be unilaterally sensitive to spatial interaction. The longer the length or size of a stimulus (physical or implicit), the longer its duration is perceived (Xuan et al., 2007; Casasanto and Boroditsky, 2008; Bottini and Casasanto, 2010; Casasanto et al., 2010; Dormal and Pesenti, 2013). In contrast, studies report that duration does not influence numerical judgments (Droit-Volet et al., 2003; Dormal et al., 2006; Dormal and Pesenti, 2013). Similarly, although the classic tau effect (Helson and King, 1931) is an example of the influence of duration on spatial judgment, this finding has often not been replicated, leading to the suggestion that duration does not influence spatial judgments (Casasanto and Boroditsky, 2008; Bottini and Casasanto, 2010; Dormal and Pesenti, 2013). An exception to this pattern

of results has been shown in a few recent studies in which interactions between number and duration have been reported to be bidirectional (Arend et al, under review; Javadi and Aichelburg, 2012, 2013). Likewise, spatial and numerical dimensions have been shown to interfere with each other bidirectionally (space affects number perception and number affects space perception), although not always symmetrically (interactions can be stronger in one direction than the other). Most studies report again a "more A, more B" pattern: the larger the numerical (symbolic or non symbolic) magnitude, the longer the length of a line is perceived (Dormal and Pesenti, 2007, 2013; De Hevia et al., 2008; De Hevia and Spelke, 2010); reciprocally, the longer the size, the larger the number is perceived (Dormal and Pesenti, 2007, 2013; although see Shuman and Spelke, 2006 and Tokita and Ishiguchi, 2011).

Interactions between dimensions have been proposed to be the side product of an automatic mapping of number, space and time on an analogue magnitude (Cantlon, 2012; Dormal and Pesenti, 2013). Alternatively interactions could be the manifestation of a statistical relationship between numerical, spatial and temporal information that we extrapolate to refine magnitude estimations (Cantlon, 2012): if we observe consistently that longer distance take a longer time and a larger number of steps to walk, we can correct our estimate of the length path by estimating the duration of the trip and the number of steps we made. In both cases, interactions reflect a tight relationship between the processing of different magnitude dimensions. To the best of our knowledge no research has yet assessed whether OAs present the same pattern of interactions at those observed in younger individuals.

OBJECTIVES OF THE CURRENT STUDIES

Here we investigated time and space processing first independently, and then in combination in young and ageing participants. In a first experiment we assessed spatial and temporal processing in OAs using a well-established psychophysics paradigm. In addition, using dedicated and well-known neuropsychological measures, we investigated the integrity of older participants' arithmetical, memory, attention and executive processes which might reflect or contribute to any age-related difference in quantity skills. We reasoned that if performance in the spatial and temporal processing tasks did not differ between older and younger participants, this may be suggestive of maintained temporal and spatial judgments in ageing, or of compensatory mechanisms in OAs to palliate to the general cognitive decline associated with ageing. Performance in tasks assessing auxiliary processes (memory, attention, executive processes) provided us with a measure of cognitive decline, allowing us to evaluate its relation to performance in time and space discrimination. In contrast, age-related differences in time and space discrimination may reflect impairments specific to a single dimension, or impairments of the whole quantity system. In a second experiment we examined the relationships between different dimensions in ageing and probed whether magnitude dimensions interfere with each other in a similar fashion in OAs and in YAs. We reasoned that a similar pattern of interactions in the two groups, albeit different in amplitude (e.g., stronger or weaker in the OAs group), may indicate that the magnitude system is robust and resilient

to ageing. Alternatively, if magnitude dimensions are differently affected by ageing, the pattern of interactions itself (i.e., directionality of the interactions) is expected to differ from YAs' without necessarily showing weaker or stronger interactions.

PARTICIPANTS

A total of 70 right-handed neurologically healthy participants with normal or corrected-to-normal vision gave written consent and were paid to participate in our study which was approved by the local research Ethics Committee. Participants were selected from the UCL Institute of Cognitive Neuroscience database based on their age. Forty-five participants took part in Study 1: 24 were young participants with a mean age of 24.8 years ($SD = 3.64$; age range 20–35; 9 males); 21 were older participants with a mean age of 65 years ($SD = 4.8$; age range 59–74; 10 males). Thirty participants took part in Study 2: 16 were young participants with a mean age of 25.0 years ($SD = 4.4$; age range: 20–37; 9 males); 14 were older participants with a mean age of 66.9 years ($SD = 3.4$; age range: 63–73; 6 males). Two young and three older participants took part in both studies.

STUDY 1

We first examined whether processing the continuous dimensions of time and space may be affected by ageing. We used an established experimental paradigm previously employed to probe continuous quantity processing in young healthy participants and neurological patients (Cappelletti et al., 2009, 2011), whereby in different blocks participants were asked to discriminate duration or spatial extension (length) on one-dimensional stimuli (horizontal lines).

METHODS
Background tasks
Participants in both groups were assessed with standard tests of intelligence (National Adult Reading Test, Nelson and Willison, 1991) and vocabulary (vocabulary subtest of the WAIS-R, Wechsler, 1995). They were also tested on the Attention Network Test (Fan et al., 2002), the color Stroop task (Stroop, 1935) and the number Stroop task (Henik and Tzelgov, 1982) to assess attentional and inhibitory functions (see description of the tasks below); the "Doors and People" test (Baddeley et al., 1994) as well as the digit span and the spatial span (Wechsler, 1995; see description of the tasks below) were administered to test memory performance. In addition, OAs were given the Mini Mental State Examination (Folstein et al., 1975) to screen for cognitive impairment.

The Attention Network Test (ANT, Fan et al., 2002) examines executive and inhibitory processes by asking participants to attend to one target while ignoring others (Posner et al., 1980). Three aspects of performance are measured: alertness, orienting, and conflict. The version used here combined a cueing task and a flanker task (Eriksen and Eriksen, 1974): participants responded to cued or un-cued central targets while ignoring flanking distractors. The stimuli consisted of a target arrow flanked by two arrows on either side, which could point to the same direction as the target arrow (congruent condition, e.g.,→→→→→) or to the opposite direction (incongruent condition, e.g.,→→←→→).

Following Fan et al. (2002), each arrow was presented at 0.55° of visual angle and separated from the adjacent arrows by 0.06° of visual angle. The stimuli (central arrow and flankers) measured 3.08° of visual angle in total. Participants were instructed to attend to the middle arrow and to decide whether it was pointing to the left or to the right. Each trial started with a central fixation cross which was presented for a random duration between 400 and 1600 ms, followed by either a 100 ms warning asterisk cue (cued trials) or by a longer fixation (un-cued trials), and by a second 400 ms fixation period after which the target and the flankers appeared simultaneously and centrally at 1.06° of visual angle either above or below the fixation point. The cue was always valid and could either appear centrally, i.e., in a spatially neutral condition or precede the target and flankers in the same position above or below the fixation point, i.e., in a spatially-orienting condition. The target and flankers remained on the screen until the participant responded or for a maximum of 1700 ms. The next trial began immediately after a response was made. A total of 288 trials were presented in 3 blocks of 96 trials each. Responses were made by pressing a left-hand key (or right-hand key) if the central arrow pointed left (or right) as quickly as possible.

The color Stroop task (Stroop, 1935) provides a standard measure of participants' ability to inhibit task-irrelevant information. Participants are instructed to report as quickly as possible the color of the font in which words are displayed while ignoring their meaning. In each trial, participants saw a centrally presented 500 ms fixation cross, followed by a word stimulus which stayed on the screen until the participant made a response or for a maximum of 4000 ms. The following trial started immediately. The task consisted of a total of 60 trials. Stimuli were either the words "RED" and "BLUE" or a string of "XXX." The color of the font was red or blue, resulting in congruent (e.g., the word RED appearing in red), incongruent (e.g., the word RED appearing in blue) and neutral conditions (e.g., XXX appearing in red). There were 20 trials in each condition. Responses were given by pressing the left or right arrow keys for blue or red color of the font, respectively.

The number Stroop task (Henik and Tzelgov, 1982) assesses the automatic processing of numbers as well as inhibitory processes using stimuli that contain congruent and incongruent information. In two separate tasks, participants viewed a total of 336 pairs of 1–9 Arabic numbers (168 per block) that could vary in numerical magnitude (e.g., 3 vs. 2) or physical size (e.g., 3 vs. 2). There were three types of stimuli (36 trials for each type): a pair in which the digit larger in magnitude was also larger in size was a congruent stimulus; a pair in which digits did not differ in one of the two dimensions was a neutral stimulus; a pair in which the digit larger in magnitude was smaller in size was an incongruent stimulus. Each number stimulus could be paired to itself, therefore consisting of a neutral stimulus for the physical size condition (e.g., 2 vs. 2), or to another number stimulus which could be between 1 and 4 units apart. Moreover, the two number stimuli could be of the same physical size, therefore consisting of the neutral stimulus for the numerical magnitude condition, or they could vary along two levels of physical size, as stimuli could appear in a vertical visual angle of 0.7 or 0.9° centered along the horizontal line of the computer screen to the left or the right of the fixation cross. Participants indicated

on which side was the larger number in either numerical magnitude or physical size by pressing either the left or the right arrow. A trial started with a 500 ms fixation cross, followed by the number stimuli until the participant made an answer or for a maximum of 4000 ms. After this, the following trial started immediately. For each task (number or physical size), accuracy and response times were recorded. This experimental paradigm commonly shows a "facilitation effect," i.e., participants are faster to respond to congruent stimuli (e.g., 3 vs. 2) relative to neutral stimuli (e.g., 3 vs. 3 for physical comparisons or 3 vs. 2 for numerical comparisons), they are slower to respond to incongruent stimuli (e.g., 3 vs. 2) relative to neutral stimuli (Henik and Tzelgov, 1982).

The "Doors and People" Recognition test ("Doors" stimuli only) was used to assess visual memory (Baddeley et al., 1994). Participants were asked to memorise the images of two sets consisting of 12 pictures of doors, which were presented sequentially for 3 s each. Immediately after, participants were asked to indicate with no time pressure which image they had previously seen amongst a choice of 4 images, three of which were new. In the first set of pictures, new and old door stimuli differed on general appearance; in the second set, old and new door stimuli differed in finer details (more difficult).

The digit span task (Wechsler, 1995) was used to assess verbal working memory. Here participants were instructed to repeat increasingly longer sequences of number stimuli presented verbally. The sequences increased in length by one item until a participant could not repeat two sequences of the same length without making an error. In a first block, the sequences had to be repeated in the forward order; in a second block, they were repeated in the reversed order.

Spatial span assessed spatial working memory using the "Corsi" task (Wechsler, 1995). Participants observed the experimenter touching a series of blocks on a horizontal board in a given sequence. Participants were then instructed to repeat the same steps in each sequence. Sequences increased progressively in length by one unit and were repeated in the forward order only.

Experimental tasks: continuous quantity processing

Stimulus presentation and data collection were controlled using the Cogent Graphics toolbox (http://www.vislab.ucl.ac.uk/Cogent) and Matlab 7.0 software on a Sony-Vaio laptop computer. The dimensions of the display, as rendered on the built-in liquid-crystal screen, were 33.8 cm horizontal by 27 cm vertical. The display had a resolution of 1280 × 1024 pixels and was refreshed at a frequency of 60 Hz. A chin-rest was used to stabilize head position of the participants and the viewing distance from the monitor was 50 cm. During all testing sessions participants sat in a quiet room facing the computer screen under normal room lighting.

Stimuli. Stimuli were two horizontal white lines (thickness 0.153°) centered on the vertical meridian on a black background. The lines were presented sequentially in a two-interval discrimination paradigm, one line 5.07° above the horizontal meridian and the other 5.07° below (see **Figure 1A**). The first line stimulus in the two-interval sequence (the "Reference") always had a
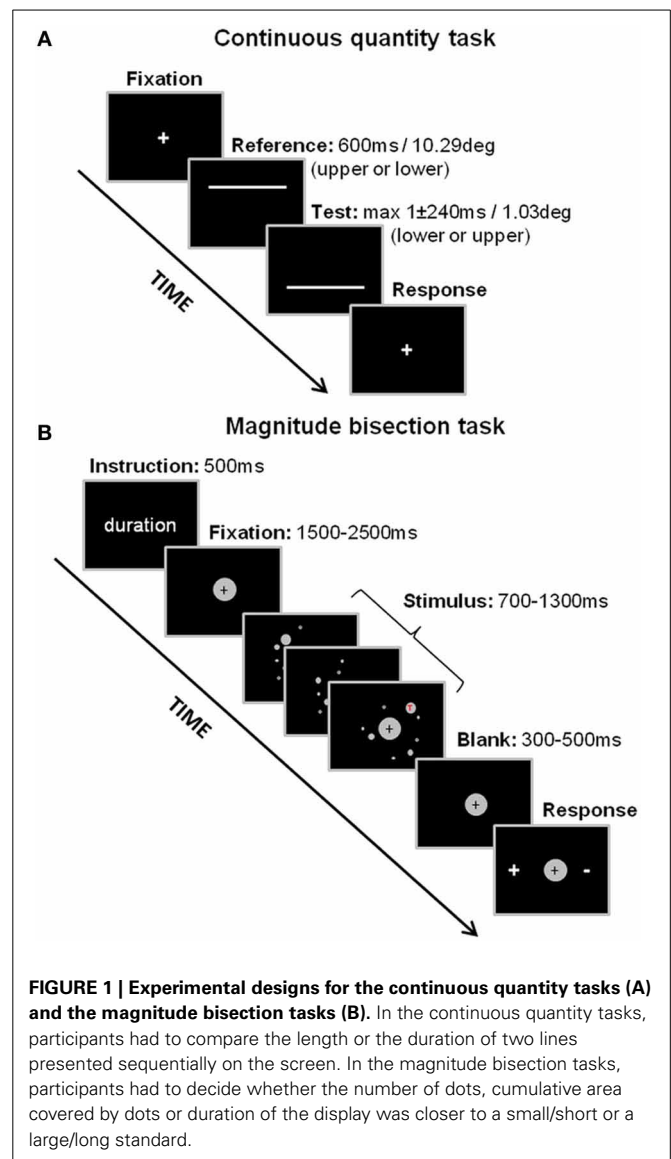
FIGURE 1 | Experimental designs for the continuous quantity tasks (A) and the magnitude bisection tasks (B). In the continuous quantity tasks, participants had to compare the length or the duration of two lines presented sequentially on the screen. In the magnitude bisection tasks, participants had to decide whether the number of dots, cumulative area covered by dots or duration of the display was closer to a small/short or a large/long standard.

length of 10.06° and a duration of 600 ms. The second line (the "Test") could vary according to a Method of Constant Stimulip either in length or duration, depending on the dimension to be judged (the irrelevant dimension always matched the Reference). For each dimension the ratio between the smaller and the larger stimulus could vary unpredictably over five levels (steps of 0.201° for length and 40 ms for time) with equal frequency: ratio of 1.02, 1.04, 1.06, 1.08, and 1.10 for length and ratio of 1.067, 1.133, 1.20, 1.267, 1.333 for time, selected from previous pilot studies. There were 5 blocks of 40 observations for each level of the test stimulus (total of 200 observations for each task). The length and duration discrimination tasks were run independently from each other in counterbalanced order across participants to avoid order effects (see **Figure 1A**).

Design. Each trial began with a centrally displayed fixation point (diameter 0.153°), which remained visible until a key-press from the participant. The Reference (or the Test) line was then

immediately displayed above (or below) the fixation point followed by the Test (or the Reference) line below (or above) the fixation, and an inter-stimulus-interval of 100ms. The screen then remained blank until a response from the subject, followed by the central fixation point which stayed on the screen until the participant pressed the space bar; the next trial followed immediately (see **Figure 1A**).

Procedure. In each task, participants were instructed to make un-speeded responses by pressing either the "up" or "down" cursor-arrow keys of the computer keyboard to indicate the vertical position of the test line which appeared the longest either in duration or in spatial extent. Correct answers were equally assigned to the "up" or "down" keys in each task.

DATA ANALYSIS

For both *color and number Stroop tasks* we calculated the difference in RT between congruent and incongruent trials, considered to be a standard measure of participants' ability to inhibit task-irrelevant information (Stroop, 1935). In further analyses we refer to this index as the Stroop effect.

In the *duration and length discrimination tasks*, each participant's response distributions were used to estimate the precision of the underlying magnitude representation, expressed as the Weber fraction (*wf*). The magnitude representations were assumed to be Gaussians with standard deviations linearly related to their means. The *wf* determines the variation of the standard deviation of the Gaussian random variables in each magnitude. The estimates of *wf* were obtained by fitting cumulative Gaussian function with log-transformed test magnitude as a predictor to the proportions of correct responses for each test level. The data were fitted using maximum likelihood criterion. The fitting function had the standard deviation as the only free parameter. The mean of cumulative Gaussian was fixed at the magnitude of Reference along relevant dimension. The standard deviations of the fitted functions were then divided by the square root of 2 to obtain the estimates of individual *wf*. A larger *wf* implies a larger overlap between two magnitude representations leading to a lower discriminability and a higher rate of incorrect responses. Therefore, a large *wf* indicates a worse performance in the task.

In the *ANT* three indexes of performance were measured based on how response times of correct answers are influenced by alerting cues, spatial cues, and flankers: alertness (cued vs. un-cued trials), orienting (central cue vs. spatial cue), and conflict (congruent vs. incongruent trials averaged across cued and un-cued, and central vs. spatial cue).

In all tasks, data were analyzed using ANOVA and *t*-tests with a *p*-value <0.05 considered significant for all statistical analyses. For the standardized tasks (with the exclusion of IQ, digit span, spatial span, Doors and People task and vocabulary) non-parametric tests were used (Mann–Whitney *U*-test).

In Study 1, a total of 4.84 and 4.86% of the data were missing in the older and in the younger group, respectively. The data sets were completed using expectation-maximization protocol implemented in the SPSS package. Only one data point was missing

in the time discrimination in the older group and none in the young group. There were no missing data points for the space discrimination task.

RESULTS
Background tasks
There was a significant group difference in IQ and vocabulary scores, with older participants outperforming the young [$t_{(43)} = 3.83$, $p < 0.001$, $p = 0.005$, and z-score approximation $= 2.78$, respectively], consistent with results reported in previous studies (Hedden and Gabrieli, 2004). Performance on the Mini Mental State Examination (Folstein et al., 1975) showed no signs of cognitive deterioration in older participants (see **Table 1**).

Older participants performed worse than young in tests assessing attention (ANT); specifically they were significantly slower at orienting [$t_{(43)} = 2.07$, $p < 0.044$] and alerting attention following visual cues [$t_{(43)} = 2.33$, $p = 0.025$]. Older participants were also worse than younger participants in processing stimuli containing conflicting information [$t_{(43)} = 3.52$, $p = 0.001$].

Executive functions measured in the color Stroop task also indicated group differences. Both groups showed reliable Stroop effect [YA: $t_{(23)} = 3.38$ $p < 0.005$; OA: $t_{(20)} = 5.68$, $p < 0.001$] but this was stronger in the older group (YA mean RT difference: 21.5 ms, $SD = 31.1$; OA mean RT difference: 111.5 ms, $SD = 89.9$, $t_{(43)} = 4.61$, $p < 0.001$), indicating a difficulty for OAs to inhibit task-irrelevant information.

In the number Stroop task, an ANOVA with task (number and physical size comparison) and group (older and younger) factors showed a main effect of task [$F_{(1, 43)} = 9.78$, $p = 0.003$, $\eta_p^2 = 0.23$] and of group [$F_{(1, 43)} = 13.14$, $p = 0.001$, $\eta_p^2 = 0.31$] but no significant interaction. Across the groups, the effect of physical size on numerical comparison was stronger than vice versa (mean Stroop effect for size-relevant task: 59.5 ms, $SD = 37.8$; mean Stroop effect for number-relevant task: 87.93 ms, $SD = 50.13$). Across tasks, there was a greater Stroop effect in the older than in the younger group (mean RT difference for YA: 59.7 ms, $SD = 21.5$; mean RT difference for OA: 89.72 ms, $SD = 33.44$), consistent with the result of the color Stroop task.

Visual memory function measured with the "Doors and People" task showed a significant group difference indicating a better performance in the younger group (younger vs. older: $p < 0.001$, z-score approximation $= 3.99$). A marginally significant group difference was observed in the task measuring spatial span (YA vs. OA: $p < 0.074$, $z = 1.79$) but not digit span (YA vs. OA: $p < 0.79$, $z = 0.26$).

Experimental tasks: continuous quantity processing
We first tested whether there was any group difference in any of the continuous quantity tasks. An ANOVA with the log *wf* of duration and length tasks as within-subject factor and group (younger and older) as between-subject factor showed only a significant main effect of task [$F_{(1, 43)} = 356.42$, $p < 0.001$, $\eta_p^2 = 0.89$, see **Figure 2**, left panel]. Specifically *wf* was higher in the duration task than in the length task (0.31 and 0.041, respectively), indicating that participants across groups demonstrated higher precision to judge length than duration. Further analyses

Table 1 | Demographic data and descriptive statistics for the younger adult (left) and older adult (right) groups in Study 1.

Task/information	Younger participants ($N = 24$)	Older participants ($N = 21$)
A. DEMOGRAPHIC INFORMATION		
Age	24.8 years ($SD = 3.6$)	65.0 years ($SD = 4.8$)
Gender	9 males	10 males
B. BACKGROUND		
Full IQ (NART[a])	114.3 ($SD = 11.22$)	125.7 ($SD = 8.4$)
Mini Mental State Examination[b]	nt	Median = 30 (min = 28)
Vocabulary[c]	Median = 50 (IQR = 7)	Median = 57 (IQR = 7.5)
C. ATTENTION AND EXECUTIVE FUNCTIONS AND MEMORY		
Attention network test (ANT)[d]		
Orienting	37.3 ms ($SD = 18.5$)	54.0. ms ($SD = 34.4$)
Alerting	24.9 ms ($SD = 15.9$)	3.5 ms ($SD = 41.8$)
Conflict (incongruent-congruent)	85.9 ms ($SD = 20.5$)	116.3 ms ($SD = 36.1$)
Word Stroop task (Stroop effect)[e]	21.5 ms ($SD = 31.1$)	111.5 ms ($SD = 89.9$)
Number Stroop task (Stroop effect)[f]		
Numerical comparison	68.3 ms ($SD = 36.6$)	110.3 ms ($SD = 54.8$)
Physical comparison	51.1 ms ($SD = 35.0$)	69.1 ms ($SD = 39.4$)
Visual memory (Door recognition)[g]	Median = 22 (IQR = 2)	Median = 20 (IQR = 4.5)
Verbal memory (Digit span)[a]	Median = 22 (IQR = 8)	Median = 22 (IQR = 6)
Spatial memory (Spatial span)[a]	Median = 10 (IQR = 6)	Median = 9 (IQR = 4.5)
D. CONTINUOUS QUANTITY PROCESSING		
Time discrimination (wf)	0.32 (95% CI: 0.06–1.65)	0.30 (95% CI: 0.097–0.92)
Space discrimination (wf)	0.043 (95% CI: 0.022–0.084)	0.039 (95% CI: 0.012–0.127)

[a] Nelson and Willison, 1991;

[b] Folstein et al., 1975; max score:30

[c] Wechsler, 1995;

[d] Fan et al., 2002;

[e] Stroop, 1935;

[f] Henik and Tzelgov, 1982;

[g] Baddeley et al., 1994;

SD, Standard deviation; IQR, Interquartile Range; CI, Confidence Interval; nt, not tested; ms, milliseconds; wf, Weber Fraction.

specific for each task show no group difference [Space: $t_{(43)} = 0.7$, ns; time: $t_{(53)} = 0.34$, ns].

We also used a regression analysis to investigate whether, within the older group, age may predict performance in space and time processing. An analysis based on regressing the log wf on participants' age showed no negative effect of age on performance (space: $t = 0.86$, $p = 0.40$, $R^2_{adj} = 0.0$; time: $t = 1.57$, $p = 0.13$, $R^2_{adj} = 0.07$; where a negative t-value implies decline with age).

There was a significant correlation between wf of length and duration tasks in the older group, $r = 0.60$, $p = 0.004$, but not in the younger group ($r = 0.16$, $p = 0.46$). However, the comparison of correlations using Fischer's Z transformation failed to show a significant group difference ($z = 1.64$, $p = 0.10$).

Next, we examined whether both within and across groups continuous quantity processing correlated with other cognitive abilities, especially the inhibitory ones. There was no correlation with measures of IQ, vocabulary, attention (orienting, alerting and conflict separately), spatial, visual and verbal memory across groups. However, in the older group better performance in duration and length discrimination negatively correlated with the Stroop effect measured in the color Stroop task (Time: $r = 0.66$,

$p = 0.001$; Space: $r = 0.45$, $p = 0.036$). In other words, older participants who could better resolve conflict were also better at discriminating continuous quantities. Length discrimination also negatively correlated with a measure of conflict resolution in ANT task, ($r = 0.50$, $p = 0.021$), but not with orienting and alerting. No correlation with the Stroop effect in the number Stroop task was observed.

These results suggest that time and space discrimination were maintained in ageing participants, who showed otherwise typical signs of healthy cognitive ageing in memory, attention and executive functions. However, in the OAs group only, better performance in time and space discrimination tasks was related to their better ability to resolve conflict. This could indicate that OAs rely more on inhibitory processes than YAs when discriminating length and duration, either to suppress the tendency to answer the second of two stimuli (presentation-order effect: Hellström, 1985; Masin and Fanton, 1989), or to solve the conflict between two competing choices. In addition, whereas OAs' performance in duration and length discrimination correlated with each other, YAs' performance did not relate to each other or to any of the cognitive functioning measures that we collected. This is consistent with a recent study looking at the behavioral and

FIGURE 2 | Weber Fractions (*wf*) for the continuous quantity tasks (A) and the magnitude bisection tasks (B). In the magnitude bisection tasks *wf* are presented in the three experimental conditions c_0 (Y_{mean}, Z_{mean}), c_{min} (Y_{min}, Z_{min}) and c_{max} (Y_{max}, Z_{max}). Weber Fractions are a measure of sensitivity in discrimination tasks. Error bars show standard deviations.
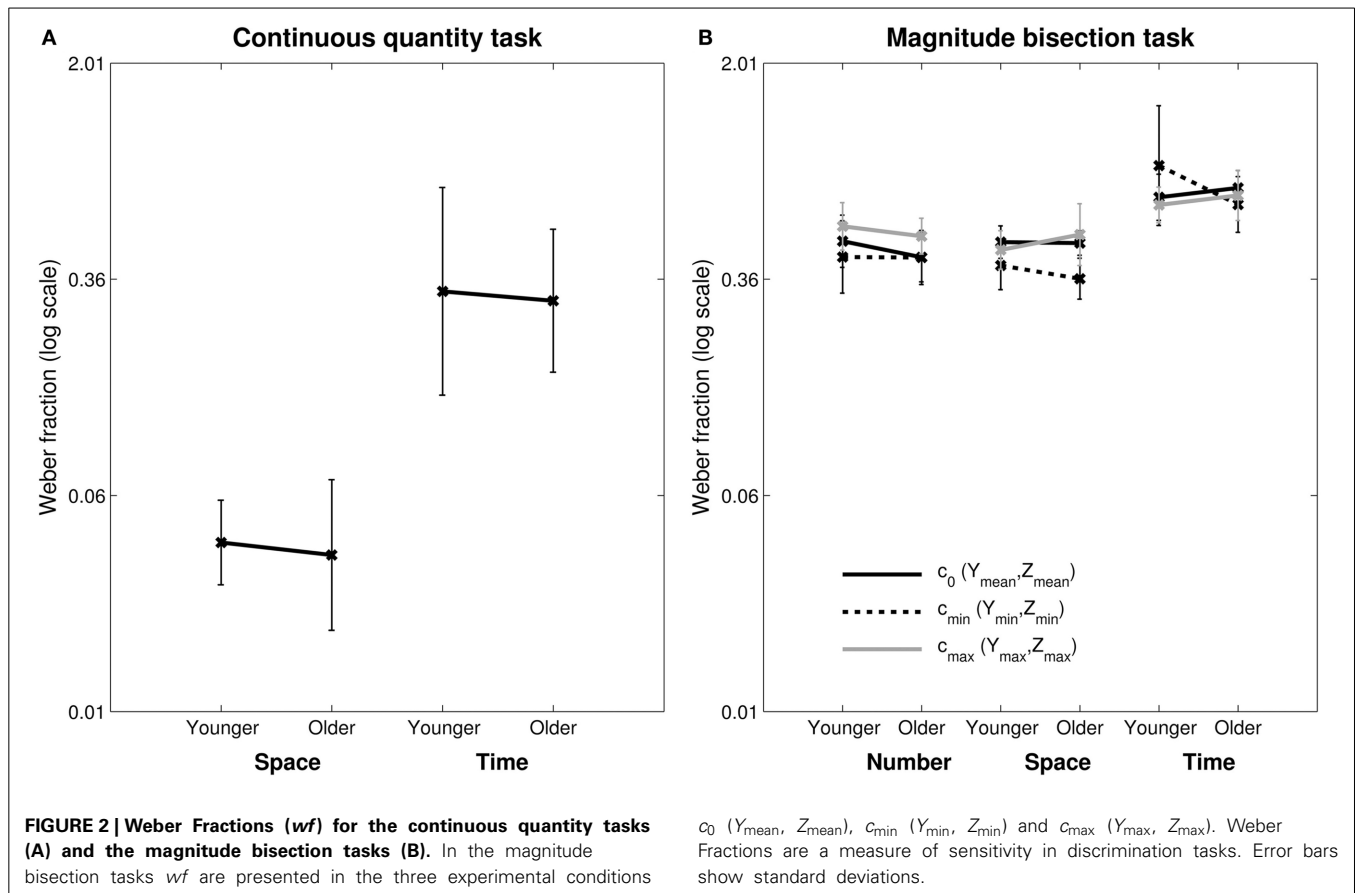
anatomical links between number and space (Cappelletti et al., in press), but it contrasts with the finding that performance in number and cumulative area discrimination correlates in young adults (Lourenco et al., 2012). The absence of correlation might indicate a stronger link between space and numerosity processing than between space and time processing. The fact that performance in duration and length discrimination tasks correlates in ageing, however, hints at the possibility that processing of continuous quantities is maintained in ageing but that the link between different dimensions might change with age. We therefore tested a second group of participants with a novel experimental design previously used to probe interactions between magnitude dimensions (Lambrechts et al., in press). We reasoned that weaker or stronger interactions between dimensions in OAs may indicate a smaller or larger reliance on common processes for magnitude discrimination, respectively. Alternatively, a pattern of interactions between dimensions which is altogether different from YAs' may suggest that age-related changes might be dimension-specific.

STUDY 2

Here we specifically examined whether known interactions between magnitude dimensions (time, space and numerosity) are maintained in ageing. The stimuli, design and procedure used were adapted from a previous paradigm employed in younger participants (Lambrechts et al., in press). Participants judged

the duration, the cumulative area covered by the stimuli, or the number of stimuli (dots) presented in a dynamic display.

METHODS

Stimulus presentation and data collection were controlled using Psychtoolbox 3.0 (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) and Matlab 7.0 software on a 1024 × 768 pixels monitor screen with a 75 Hz frame rate. Participants were seated ∼60 cm away from the display.

Background tasks

In addition to the experimental tasks participants were also assessed with a standard test of intelligence (National Adult Reading Test, Nelson and Willison, 1991) and two tests of arithmetic performance (arithmetic subtest in the WAIS-R, Wechsler, 1995 and Graded Difficulty Arithmetic test, Jackson and Warrington, 1986). The latter two were tested in order to evaluate whether a deficit in arithmetic performance may also be present, should duration and cumulative area perception be impaired.

Experimental tasks: magnitude bisection tasks

Stimuli. Stimuli were dynamic displays of gray dots which appeared and disappeared progressively within a virtual central disk on a black background on the screen (**Figure 1B**). During one display, dots appeared on the computer monitor in 5–13 steps (to produce a progressive accumulation), 1–8 dots at a time, and

they then disappeared progressively after a lifetime of 333–507 ms (all values chosen pseudo-randomly for each trial). Steps during which new dots appeared lasted 40–507 ms. A display was characterized by its duration (time elapsed between the appearance of the first dot and the disappearance of the last one), the cumulative area of its dots, and the total number of dots presented. Duration, cumulative area and number were defined according to 3 experimental conditions (see Design below). Dot stimuli had a radius comprised within 0.45 and 2.84° of visual angle, and could not overlap in space or time. The virtual disk for display had a radius varying pseudo-randomly between 5.7 and 7.7° of visual angle. Dot stimuli were constrained not to appear in an inner disk of radius 0.9° centered on a fixation cross. Luminance of all dots for one trial took one of six values [57, 64, 73, 85, 102, and 128 in the 0(black)-to-255(white) RGB-coded referential] chosen pseudo-randomly. In addition, there was a letter that appeared inside one of the dots, which could be either a red or green, upright or upside-down capital "T" (see **Figure 2A**). This was used as a control condition to test for participants' general alertness during the task (see below). Participants were asked to discriminate the target (upright red T) from the distracters (upright green T or upside-down red or green T).

Design. The experiment combined a bisection task (on magnitude dimensions) and a signal detection paradigm (target/non target). The design for both tasks is summarized in **Table 2**. For the bisection task, participants were first trained to discriminate between a small ("−") and a large ("+") standard in each magnitude dimension (short/long duration, small/large cumulative area, small/big number of dots). During the test phase they were then asked to judge whether the duration, cumulative area or number of dots in each trial was closer to the "−" standard or to the "+" standard. Standards were defined as 0.7 ("−") and 1.3 ("+") times a mean value set as $D_{mean} = 1000$ ms, $S_{mean} = 878$ mm^2, and $N_{mean} = 28$ dots for duration (D), cumulative area (S), and number (N) respectively. These values were chosen to produce similar sensitivity in the three tasks based on Lambrechts et al. (in press). During the test phase each magnitude dimension took 5 possible values defined as 0.7, 0.9, 1, 1.1, and 1.3 times the mean value (hereafter: $X_{0.7}$, $X_{0.9}$, X_{mean}, $X_{1.1}$, and $X_{1.3}$, with dimension X being D, S, or N). Three experimental conditions were retained to explore the susceptibility of the target magnitude judgment to irrelevant dimensions (see **Figure 2B**). In control *condition 0* (c_0), orthogonal dimensions were set to their mean (Y_{mean}, Z_{mean}); in *condition 1* (c_{min}), they were set to their minimal values (0.7 × mean value: Y_{min}, Z_{min}) and in *condition 2* (c_{max}), they were set to their maximal values (1.3 × mean value: Y_{max}, Z_{max}).

In addition to the magnitude tasks, we also used a target detection control task to measure participants' attention. This aimed at excluding that any generalized impairment in the magnitude tasks could be due to attention-related disorders. For this target detection task a letter appeared inside one of the dots in each trial and could be either a target (red upright T) or a distractor (red inverted T, green upright T or green inverted T).

Trials were pseudo-randomized across tasks and conditions. A total of 720 trials were collected in the magnitude bisection tasks (3 dimensions × 3 conditions × 4 values × 20 trials) and 200 in the control detection task (one third with a target and two thirds with distracters equally presented). Trials were pseudo-randomized across tasks and conditions, and blocked by 100 trials (the original experimental design comprehended two additional conditions, with a total of 1400 trials).

Table 2 | Experimental design for Study 2.

	Instruction (target magnitude)	Target magnitude value	Non-target magnitude value
Magnitude bisection task	Duration (D) Surface (S) Number (N)	$X_{min} = 0.70\,X_{mean}$ $X_{0.9} = 0.90\,X_{mean}$ $X_{1.1} = 1.10\,X_{mean}$ $X_{max} = 1.30\,X_{mean}$	If target magnitude D $c_0 = [S_{mean}, N_{mean}]$; $c_{min} = [S_{min}, N_{min}]$; $c_{max} = [S_{max}, N_{max}]$ If target magnitude S $c_0 = [D_{mean}, N_{mean}]$; $c_{min} = [D_{min}, N_{min}]$; $c_{max} = [D_{max}, N_{max}]$ If target magnitude N $c_0 = [S_{mean}, D_{mean}]$; $c_{min} = [S_{min}, D_{min}]$; $c_{max} = [S_{max}, D_{max}]$
	Instruction	**Letter stimulus**	**Letter stimulus type**
Target detection task	Red T search	Red upright T Red upside-down T Green upright T Green upside-down T	Target Distractor Distractor Distractor

*In the magnitude bisection tasks (top line), one of three dimensions (number, surface, duration) was the target dimension. Four values were tested for the target dimension X (0.7, 0.9, 1.1 and 1.3 * X_{mean}), while non-target dimensions Y and Z were determined according to three experimental conditions: in c_0 non-target dimensions took their middle value (Y_{mean}, Z_{mean}); in c_{min} non-target dimensions were minimal (Y_{min}, Z_{min}); in c_{max} non-target dimensions were maximal (Y_{max}, Z_{max}). In the target detection task (bottom line), participants had to detect a target letter (red upright T) that appeared at each trial within one of the dots and reject the distracters (green upright T, red upside-down T, green upside-down T).*

Procedure. Before the test session, participants engaged in the training phase: they were familiarized with the minimum ($X_{0.7}$) and maximum ($X_{1.3}$) values for each magnitude dimension (*D*, *S*, *N*) as well as with the target (red upright T) and distracters (green upright T, red or green upside-down T). The training session consisted of two stages: a learning and a test stage. During the learning stage, participants passively viewed 10 examples of stimuli for each task (5 minima and 5 maxima or 4 targets and 6 distracters). They then moved on to the test stage in which they were presented with the same 10 examples and asked to perform a categorical judgment. In the test phase each trial started with one of four instructions: "Duration," "Surface," "Number" (magnitude bisection task) or "Red T" (detection task) displayed centrally on the screen for 500 ms. A fixation cross followed for a duration pseudo-randomly chosen between 1500 and 2500 ms after which the stimulus was presented. After the stimulus display and a subsequent 300–500 ms fixation cross, participants were prompted for their response by the simultaneous appearance of "+" and "−" displayed on each side of the fixation cross. In the magnitude tasks (bisection tasks), participants were instructed to judge whether the stimulus displayed was closer to the minimum standard ("−") or the maximum standard ("+") in a given dimension. In the control task (target detection) participants were instructed to indicate whether they had seen either the target ("+") or a distractor ("−"), see **Figure 1B**. The relative position of "+" and "−" on the monitor was pseudo-randomly assigned throughout the trials. Response keys were "h" and "j" on the computer keyboard. Participants were instructed at the beginning to avoid counting and to respond by hunch. In addition, performance in discriminating durations in this range (700–1300 ms) is unlikely to benefit from using a counting strategy (Grondin et al., 2004). There was no time constraint to respond.

DATA ANALYSIS
Magnitude bisection tasks
The proportions of "+" responses (stimulus estimated as closer to the maximum standard) were computed separately for each task, dimension and condition. Values were individually fitted to a cumulative Gaussian function *f* using Psignifit 3.0.8 (Fründ et al., 2011) in Matlab 7.0. Two indices were computed: the Point of Subjective Equality (*PSE*, value at 50% of "+" responses) which is a canonical measure of accuracy and the *wf*, computed as in study 1, which reflect sensitivity.

Data were cleaned as follow: when *wf* values were negative or outside ±3 standard deviations of the total mean, data for that participant and in that dimension were excluded (10% of the data were excluded across both groups). In the YA group, 2 participants were excluded from the duration task, 2 from the cumulative area task and 1 from the number task. In the OA group, 2 participants were excluded from the duration task and 2 were excluded from the cumulative area task (resulting in a minimum of 12 participants per dimension in each group).

Separate repeated-measure ANOVAs were performed on PSEs and *wf*s using the IBM SPSS software (Version 19.0). A Greenhouse-Geisser correction was applied when appropriate. *Post-hoc* Bonferroni-corrected *t*-tests were performed to explore significant main effects or interactions.

Target detection task
Hit and false alarm rate were computed as the proportion of target which were correctly detected, and the proportion of distractors that were detected as targets, respectively. Dprime (*d′*) detection scores were computed by subtracting the z-scores of hit from false alarm (with *N* the inverse normal law):

$$d' = N^{-1}(\text{HIT}) - N^{-1}(\textit{FA})$$

RESULTS
Background tasks
The two groups differed marginally in the estimate of IQ assessed by the National Adult Reading Test, with OAs slightly outperforming young adults [$t_{(23)} = -2.03$, $p = 0.054$]. Participants in the two groups did not differ in arithmetic performance as measured by the Graded Difficulty Arithmetic test [$t_{(27)} = -0.065$, $p > 0.9$] and the arithmetic subtest of the Wechsler Adult Intelligence Scale-R [$t_{(27)} = -0.613$, $p > 0.5$].

Target detection task
Both groups were able to perform the target detection task ($d'_{\text{YA}} = 2.18$; $d'_{\text{OA}} = 2.18$), with no group difference [independent sample *t*-test on *d′* values, $t_{(28)} = 0.02$, $p > 0.9$]. This suggests that both groups were equally able to attend to the stimuli throughout the task.

Experimental tasks: magnitude bisection tasks
Since our criterion to include or exclude individual participant's data point was applied separately for each task, some participants were retained in one task and not in others. In order to maximize statistical power we therefore conducted statistical analyses on each task separately. **Figure 3** shows the psychometric profiles of responses obtained in each group.

Point of Subjective Equality (PSE). Planned mixed-design, repeated-measure ANOVAs with PSE as the dependent variable, condition (3: c_0, $c_{\min}$, $c_{\max}$) as independent factor and group (2: YA, OA) as between-group factor were conducted for each task separately. Results are presented in **Figure 3** (right panel).

In every task, the ANOVA revealed a significant main effect of condition [number: $F_{(2, 29)} = 138.64$, $p < 0.001$, $\eta_p^2 = 0.828$; cumulative area: $F_{(2, 26)} = 156.96$, $p < 0.001$, $\eta_p^2 = 0.867$; duration: $F_{(2, 26)} = 11.97$, $p < 0.005$, $\eta_p^2 = 0$]. In the *number task*, post-hoc paired-sample *t*-tests revealed that PSE in $c_{\min}$ was smaller than in c_0 [$t_{(29)} = 10.93$, $p < 0.001$] and PSE in c_0 was smaller than in $c_{\max}$[$t_{(29)} = -7.22$, $p < 0.001$]. Additionally, PSE was smaller in $c_{\min}$ than in $c_{\max}$[$t_{(29)} = -14.51$, $p < 0.001$]. Therefore, number was overestimated when duration and cumulative area were minimal, and underestimated when duration and cumulative area were maximal.

Similarly, in the *cumulative area task*, post-hoc paired *t*-tests showed that PSE in $c_{\min}$ was smaller than in c_0 [$t_{(26)} = 10.07$, $p < 0.001$] and PSE in c_0 was smaller than in $c_{\max}$[$t_{(26)} = -9.19$, $p < 0.001$]. Additionally, PSE was smaller in $c_{\min}$ than in $c_{\max}$[$t_{(26)} = -15.15$, $p < 0.001$]. Therefore, cumulative area was underestimated when duration and number were maximal, and overestimated when duration and number were minimal.

FIGURE 3 | Performance in the magnitude bisection tasks for the younger adult (Top) and older adult (Bottom) groups. Left panel: psychometric profiles of response for the number, cumulative area and duration tasks. Data points show average responses across participants. Full lines correspond to the average psychometric fit. **Right panel**: Point of Subjective Equality for the number, cumulative area and duration task in the three experimental conditions c_0 (Y_{mean}, Z_{mean}), c_{min} (Y_{min}, Z_{min}) and c_{max} (Y_{max}, Z_{max}). Error bars show standard error of the mean.

In the *duration task*, *post-hoc* paired-sample *t*-tests showed that PSE in both c_0 [$t_{(26)} = -3.57$, $p < 0.005$] and c_{max} [$t_{(26)} = 3.70$, $p < 0.005$] conditions were smaller than PSE in c_{min}. Duration was therefore underestimated when cumulative area and number were minimal compared to when they had either mean or maximal values.

Critically, a main effect of group was found in the *duration task* [$F_{(1, 26)} = 6.70$, $p < 0.05$, $\eta_p^2 = 0.218$]. YAs produced a higher PSE than OAs, i.e., OAs overestimated duration relative to YAs ($PSE_{YA} = 1.206$, $PSE_{OA} = 1.038$). This results confirms the idea that temporal estimation changes with age (time seems to stretch for longer). However, the absence of a condition × group interaction indicates that although the absolute perception of duration changes with ageing, the way in which other magnitude dimensions interfere with duration judgment is similar in both age groups.

Overall these results confirm that even when they are task-irrelevant, magnitude dimensions interfere with the accuracy of each other's judgment. In line with previous findings (e.g., Xuan et al., 2007, 2009; Casasanto and Boroditsky, 2008; Oliveri et al., 2008), cumulative area and numerosity affected duration judgment in a positively correlated way (the larger the cumulative area and number, the longer the subjective duration). More surprisingly, duration and number, and duration and cumulative area, affected cumulative area and number judgment in a negatively correlated way, respectively; this means that many dots presented for a longer time appeared to have a small cumulative area, and that larger dots presented for a longer time seemed less numerous.

Additionally, while cumulative area and number were perceived similarly by both groups, temporal content was judged shorter in the older than in the YAs group.

Weber fraction (wf). Planned mixed-design, repeated-measure ANOVAs with *wf* as the dependent variable, condition (3: c_0, c_{min}, c_{max}) as independent factor and group (2: YA, OA) as between-group factor were run for each task separately. Results are presented in **Figure 2** (right panel).

In the *number task*, the ANOVA revealed a main effect of condition [$F_{(2, 29)} = 10.31$, $p < 0.001$, $\eta_p^2 = 0.276$]. Corrected *post-hoc* paired-sample *t*-tests revealed that *wf* was lower in c_{min} [$t_{(29)} = -3.88$, $p < 0.005$] and in c_0 [$t_{(29)} = -2.92$, $p < 0.01$] than in c_{max}. *Wf* was also found to be marginally smaller in c_{min} than in c_0 [$t_{(29)} = 1.94$, $p = 0.062$]. This indicates that both groups were less precise to estimate number when space and time had large values than when they had small values. Critically, no main effect or interaction with group was significant, indicating that older and YAs estimated number equally well.

In the *cumulative area task*, the ANOVA revealed a main effect of condition [$F_{(2, 26)} = 13.10$, $p < 0.001$, $\eta_p^2 = 0.353$]. Corrected *post-hoc* paired-sample *t*-tests showed that *wf* was lower in c_{min} than both in c_0 [$t_{(26)} = 6.51$, $p < 0.001$] and c_{max} [$t_{(26)} = -3.67$, $p < 0.005$], suggesting that both YAs and OAs were more precise to estimate cumulative area when time and number were minimal. Interestingly, the analysis also revealed a significant condition × group interaction [$F_{(2, 26)} = 4.59$, $p < 0.05$, $\eta_p^2 = 0.161$]. *Post-hoc* independent *t*-tests, however, showed no significant

difference between YAs and OAs' *wf* in either of the conditions (c_0, c_{min}, or c_{max}). Further independent *t*-tests revealed that the interaction was likely driven by the difference between *wf* in c_{min} and c_{max} ($wf_{c2} - wf_{c1}$) which differed in YAs and OAs [$t_{(24)} = -2.51$, $p < 0.05$]. Paired *t*-tests indicated that in both groups *wf* were smaller in c_{min} than both in c_{max} [YA: $t_{(14)} = -2.84$, $p < 0.05$; OA: $t_{(12)} = -3.29$, $p < 0.01$] and c_0 [YA: $t_{(14)} = 3.86$, $p < 0.005$; OA: $t_{(12)} = 5.58$, $p < 0.001$]. Participants were more precise to judge cumulative area when few dots were presented for a short duration (in c_{min}) than when many dots were presented for a long duration (in c_{max}), and even more so in the OAs than in the YAs group. There was no significant main effect of group, confirming that precision in the cumulative area task was very similar in both participants group, although interaction effects were slightly accentuated in the OAs group.

In the *duration task*, the ANOVA revealed a marginal main effect of condition [$F_{(2, 26)} = 3.37$, $p = 0.076$, $\eta_p^2 = 0.123$] and a significant interaction condition × group [$F_{(2, 26)} = 4.87$, $p < 0.05$, $\eta_p^2 = 0.169$]. *Post-hoc* independent *t*-tests revealed that *wf* in c_{min} was larger in the YAs than in the OAs group [$t_{(24)} = 2.01$, $p < 0.05$]. Paired *t*-tests further indicated that in the YAs group *wf* was smaller in c_{max} and as an index than in c_{min} [$t_{(14)} = 2.33$, $p < 0.05$ and $t_{(14)} = -2.15$, $p = 0.051$, respectively] whereas there were no difference between conditions in the OAs group. This indicates that sensitivity for duration increased when cumulative area and number took larger values in the YA group, whereas sensitivity to duration was unaffected by cumulative area and number in the OAs group. There was no significant main effect of group, indicating that overall precision in the duration task was similar in both groups.

Overall our findings indicated that sensitivity to number judgment was modulated by task-irrelevant dimensions similarly in both groups. In contrast, in the cumulative area and duration tasks, the fine pattern of interactions differed between groups. In the cumulative area task, OAs' performance was more sensitive to interaction than that of YAs, whereas in the duration task OAs' performance was more resilient to interaction than that of YAs. However, *wf* did not overall differ between age groups in any of the dimensions, suggesting that overall quantity discrimination is preserved in ageing.

DISCUSSION

This research evaluated the integrity of quantity processing in healthy ageing. In the first study, we used a two-choice paradigm to investigate continuous quantity processing (space and time discrimination) along with standard measures of cognitive processing. Our results indicate that although elderly participants showed typical age-related decline in memory, attention and executive functions, the ability to judge space and time remained intact. To further assess quantity processing in ageing, we explored the relation between magnitude dimensions whose interactions, so far observed in childhood and in young adulthood, have been taken to suggest the existence of shared or overlapping resources for quantity processing (Gallistel and Gelman, 2000; Walsh, 2003; Bueti and Walsh, 2009; Cantlon, 2012). In a second study, we therefore tested whether processing of number, time and space also interacted with each other in

older as well as in younger participants. We found that irrespective of age, number, duration and cumulative area estimations were susceptible to concurrent, task-irrelevant magnitude dimensions, suggesting that quantity processing may be supported by a shared mechanism throughout adulthood. However, the extent to which task-irrelevant dimensions influence the sensitivity of continuous quantity judgments (cumulative area and duration) differed slightly with age. In addition, the percept of duration was found to be modulated by age as elderly adults judged durations close to their veridical values whereas younger adults tended to underestimate duration.

Our results of preserved continuous quantity processing (space and time) in ageing, despite otherwise typical signs of cognitive decline, is to our knowledge the first evidence of the integrity of continuous quantity discrimination in healthy ageing. Together with recent findings showing that numerosity discrimination is also resilient to age (Cappelletti et al., in press), this suggests that non-symbolic quantity processing is generally preserved in healthy ageing. This finding might appear in contrast to other studies showing that flexibility in arithmetical problem solving tasks (e.g., Geary and Lin, 1998; Duverne and Lemaire, 2005; Lemaire and Arnaud, 2008) and performance in temporal estimation tasks (e.g., Block et al., 1998; Baudouin et al., 2006; Lustig and Meck, 2011) decrease with age. However, past research has pointed out that decline in other cognitive functions and processes such as memory, processing speed, attention or executive functions rather than quantity processing itself might account for reduced performance in some numerical and temporal judgment tasks (Salthouse, 1991; Salthouse and Kersten, 1993; Vanneste and Pouthas, 1999; Perbal et al., 2002; Salthouse et al., 2003; Duverne et al., 2008; Cappelletti et al., in press).

Our evidence of maintained quantity processing adds to other cognitive abilities that have proven resilient to ageing, such as verbal memory (vocabulary), implicit memory and emotional processes (Hedden and Gabrieli, 2004), and as such our finding contributes to defining the profile of preserved and declining cognitive abilities in older age (Hedden and Gabrieli, 2004). At present, it is not clear why some cognitive processes are better preserved than others in ageing. One possibility is that quantity-based processes may be more primitive and therefore more robust than later acquired skills such as arithmetic or second-language acquisition. Although quantity processes refine with age, they are in place very early in development (e.g., Xu and Spelke, 2000; Feigenson et al., 2002; Brannon et al., 2007). Their ubiquity makes them crucial to navigate the environment at any age. Preserving them in ageing, either by maintaining the same strategies or by reallocating resources could allow individuals to remain aware of their environment and able to adapt their behavior accordingly.

We also found that OAs showed patterns of interaction among quantities which resemble those observed in children and young adults and which have led to the hypothesis of a common mechanism for time, space and number processing (Walsh, 2003; Bueti and Walsh, 2009; Cantlon, 2012). Although most studies have postulated that interactions result from the automatic mapping of quantities onto a unique mental representation (Henik and Tzelgov, 1982; Dehaene, 1992; Dormal et al., 2006; De Hevia and Spelke, 2010; Chang et al., 2011), others (e.g., Lambrechts et al.,

in press) proposed that quantity estimates more likely result from Bayesian-like cue-integration whereby the preferred strategy to estimate quantity is to combine cues not only from the target dimension but also from concurrent dimensions. A similar view was expressed in Karolis (2013) and supported by an analysis of the scales for space and number. Here, we found that interactions related to cumulative area and duration (as observed on a measure of sensitivity) were modulated by age. For instance, when judging cumulative area, OAs were more susceptible to task-irrelevant magnitude information than YAs. In contrast, when judging durations, OAs were more resilient to interaction of other magnitude dimensions than YAs. Such observations are difficult to reconcile with the view of aging as a declining evolution. For instance, the Inhibition Deficit theory (Hasher and Zacks, 1988) claims that the ability to inhibit task-irrelevant information decreases with age and would predict that interactions are amplified in ageing. However, this would only account for the group differences obtained in the cumulative area task and not in the duration task. A more parsimonious interpretation would be that the weight with which each dimension affects the others changes with age. The current design and our relatively small sample size in study 2 do not allow us to conclude on this possibility which should be explored in the future using dedicated paradigms.

Interestingly, irrespective of age, the directions of the interactions we observed were different from those often reported in the literature. While space and number positively interacted with time perception (more, larger dots were judged to last longer) similar to previous studies (e.g., Dormal et al., 2006; Xuan et al., 2007, 2009; Casasanto and Boroditsky, 2008; Oliveri et al., 2008; Chang et al., 2011), space and time negatively interacted with number estimates, and number and time negatively interacted with space estimates. For instance, larger dots presented for a longer time were estimated less numerous and more dots presented for a longer time were estimated as covering a smaller space. Previous studies reported the opposite pattern, namely that concurrent quantities positively interact with each other (e.g., Pinel et al., 2004; Dormal and Pesenti, 2007; Javadi and Aichelburg, 2012). These unpredicted results, which replicate recent findings obtained with a similar paradigm (Lambrechts et al., in press), may be explained by differences in the experimental paradigm used here and in past studies. Crucially, in our paradigm information about all three quantity dimensions was designed to accumulate similarly over time to match the intrinsic continuous property of duration. Therefore, participants had to integrate time, space and number over the course of the stimuli presentation and could not access the total cumulative area or total number of dots at any single time point before the end. As a result, the stimulus duration affected the amount of dots presented on the screen at a given time. For instance, given the same number of dots, when the stimulus duration was longer (or shorter), less dots were presented on average at a given moment, which could lead participants to perceive them as less (or more) numerous than veridical, arguably misleading them into underestimating (or overestimating) their number. This contrasts with previous studies in which spatial and numerical information were usually displayed all at once on the screen and stayed for the whole duration of the stimulus presentation (e.g., Xuan et al., 2007,

2009; Oliveri et al., 2008; Chang et al., 2011, but see Casasanto and Boroditsky, 2008). In these studies, participants could estimate space and number as soon as a stimulus was presented, independently from its duration, so time did not impact numerical and spatial processing.

Another unexpected result was that older participants produced smaller PSE than younger participants in the duration task in Study 2 estimates of duration were closest to the veridical value for older than YAs. This finding is in disagreement with past research on time perception in ageing claiming that the ratio of estimated duration on objective duration increases with age, i.e., PSE should be getting larger with age (Block et al., 1998). It should be pointed out that most studies used different paradigms such as duration production and reproduction; importantly they tested longer durations (a few seconds or more) than the ones assessed in the present study. The study by Lustig and Meck (2011) comes closest to the present methodology by using a bisection task with durations ranging from 3–6 s, and reports—similar to previous studies—that OAs produce a larger PSE than YAs. Based on time perception models, differences between YAs and OAs were interpreted by most authors in terms of decreased attentional span in the older participant group, although attentional skills were not directly assessed in these studies. Instead in our study, we controlled for attentional levels which were very similar in both groups. In addition, the use of shorter durations might have attenuated the load on attentional processes to maintain information throughout a trial.

CONCLUSION

Here we examined the integrity of continuous quantity processing and the link between number, space and time in ageing. We showed first that discrimination of space and time, much like number, was preserved in ageing. We argued that the resilience of quantity processing skills in ageing may reflect the stability of primitive resources dedicated to quantity processing. Second, extending previous findings obtained with children and young adults, we demonstrated that in older adults, number, space and time interact in discrimination judgments, similar to what is observed in younger participants. However, we found subtle dimension-specific differences in the way concurrent dimensions affected the precision of continuous quantity estimation between younger and older adults which might indicate a change of weight of each dimension within the magnitude processing system.

REFERENCES

Baddeley, A. D., Emslie, H., and Nimmo-Smith, I. (1994). *Doors and People: a Test of Visual and Verbal Recall and Recognition*. Bury St. Edmunds: Thames Valley Test Company.

Basso, G., Nichelli, P., Frassinetti, F., and Di Pellegrino, G. (1996). Time perception in a neglected space. *Neuroreport* 7, 2111–2114. doi: 10.1097/00001756-199609020-00009

Baudouin, A., Vanneste, S., Pouthas, V., and Isingrini, M. (2006). Age-related changes in duration reproduction: involvement of working memory processes. *Brain Cogn.* 62, 17–23. doi: 10.1016/j.bandc.2006.03.003

Beran, M. J. (2007). Rhesus monkeys (*Macaca mulatta*) enumerate large and small sequentially presented sets of items using analog numerical representations. *J. Exp. Psychol. Anim. Behave. Process*. 33, 42–54. doi: 10.1037/0097-7403.33.1.42

Birren, J. E., and Botwinick, J. (1955). Speed of response as a function of perceptual difficulty and age. *J. Gerontol*. 10, 433–436. doi: 10.1093/geronj/10.4.433

Block, R. A., Zakay, D., and Hancock, P. A. (1998). Human aging and duration judgments: a meta-analytic review. *Psychol. Aging* 13, 584–596. doi: 10.1037/0882-7974.13.4.584

Bonn, C. D., and Cantlon, J. F. (2012). The origins and structure of quantitative concepts. *Cogn. Neuropsychol*. 29, 149–173. doi: 10.1080/02643294.2012.707122

Bottini, R., and Casasanto, D. (2010). "Implicit spatial length modulates time estimates, but not vice versa," in *Processings of the 32nd Annual Conference of the Cognitive Science Society*, (Mt. Hood/Portland, OR), 1348–1353.

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis*. 10, 433–436. doi: 10.1163/156856897X00357

Brannon, E. M., Lutz, D., and Cordes, S. (2006). Number representation in infancy. *Dev. Sci*. 9, 59–64. doi: 10.1111/j.1467-7687.2006.00530.x

Brannon, E. M., Suanda, S., and Libertus, K. (2007). Temporal discrimination increases in precision over development and parallels the development of numerosity discrimination. *Dev. Sci*. 10, 770–777. doi: 10.1111/j.1467-7687.2007.00635.x

Breukelaar, J. W., and Dalrymple-Alford, J. C. (1998). Timing ability and numerical competence in rats. *J. Exp. Psychol. Anim. Behave. Process*. 24, 84–97. doi: 10.1037/0097-7403.24.1.84

Bueti, D., and Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Phil. Trans. R Soc. B Biol. Sci*. 364, 1831–1840. doi: 10.1098/rstb.2009.0028

Cantlon, J. F. (2012). Math, monkeys, and the developing brain. *Proc. Natl. Acad. Sci. U.S.A*. 109(Suppl.), 10725–10732. doi: 10.1073/pnas.1201893109

Cappelletti, M., Didino, D., Stoianov, I., and Zorzi, M. (in press). Number skills are maintained in healthy ageing. *Cogn. Psychol*.

Cappelletti, M., Freeman, E. D., and Cipolotti, L. (2009). Dissociations and interactions between time, numerosity and space processing. *Neuropsychologia* 47, 2732–2748. doi: 10.1016/j.neuropsychologia.2009.05.024

Cappelletti, M., Freeman, E. D., and Cipolotti, L. (2011). Numbers and time doubly dissociate. *Neuropsychologia*. 49, 3078–3092. doi: 10.1016/j.neuropsychologia.2011.07.014

Casasanto, D., and Boroditsky, L. (2008). Time in the mind: using space to think about time. *Cognition* 106, 579–593. doi: 10.1016/j.cognition.2007.03.004

Casasanto, D., Fotakopoulou, O., and Boroditsky, L. (2010). Space and time in the child's mind: evidence for a cross-dimensional asymmetry. *Cogn. Sci*. 34, 387–405. doi: 10.1111/j.1551-6709.2010.01094.x

Chang, A. Y.-C., Tzeng, O. J. L., Hung, D. L., and Wu, D. H. (2011). Big time is not always long: numerical magnitude automatically affects time reproduction. *Psychol. Sci*. 22, 1567–1573. doi: 10.1177/0956797611418837

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N

De Hevia, M.-D., Girelli, L., Bricolo, E., and Vallar, G. (2008). The representational space of numerical magnitude: illusions of length. *Q. J. Exp. Psychol*. 61, 1496–1514. doi: 10.1080/17470210701560674

De Hevia, M. D., and Spelke, E. S. (2010). Number-space mapping in human infants. *Psychol. Sci*. 21, 653–660. doi: 10.1177/0956797610366091

Doricchi, F., Guariglia, P., Gasparini, M., and Tomaiuolo, F. (2005). Dissociation between physical and mental number line bisection in right hemisphere brain damage. *Nat. Neurosci*. 8, 1663–1665. doi: 10.1038/nn1563

Dormal, V., Grade, S., Mormont, E., and Pesenti, M. (2012). Dissociation between numerosity and duration processing in ageing and early Parkinson's disease. *Neuropsychologia* 50, 2365–2370. doi: 10.1016/j.neuropsychologia.2012.06.006

Dormal, V., and Pesenti, M. (2007). Numerosity-length interference. *Exp. Psychol*. 54, 289–297. doi: 10.1027/1618-3169.54.4.289

Dormal, V., and Pesenti, M. (2013). Processing numerosity, length and duration in a three-dimensional Stroop-like task: towards a gradient of processing automaticity? *Psychol. Res*. 77, 116–127. doi: 10.1007/s00426-012-0414-3

Dormal, V., Seron, X., and Pesenti, M. (2006). Numerosity-duration interaction: a Stroop experiment. *Acta Psychol*. 121, 109–124. doi: 10.1016/j.actpsy.2005.06.003

Droit-volet, S., Clément, A., and Fayol, M. (2008). Time, number and length?: similarities and differences in discrimination in adults and children. *Exp. Psychol*. 61, 1827–1846. doi: 10.1080/17470210701743643

Droit-Volet, S., Clément, A., and Fayol, M. (2003). Time and number discrimination in a bisection task with a sequence of stimuli: a developmental approach. *J. Exp. Child Psychol*. 84, 63–76. doi: 10.1016/S0022-0965(02)00180-7

Duverne, S., and Lemaire, P. (2005). Arithmetic split effects reflect strategy selection: an adult age comparative study in addition comparison and verification tasks. *Can. J. Exp. Psychol*. 59, 262–278. doi: 10.1037/h0087479

Duverne, S., Lemaire, P., and Vandierendonck, A. (2008). Do working-memory executive components mediate the effects of age on strategy selection or on strategy execution. Insights from arithmetic problem solving. *Psychol. Res*. 72, 27–38. doi: 10.1007/s00426-006-0071-5

El Yagoubi, R., Lemaire, P., and Besson, M. (2005). Effects of ageing on arithmetic problem-solving: an event-related brain potential study. *J. Cogn. Neurosci*. 17, 37–50. doi: 10.1162/0898929052880084

Eriksen, B. A., and Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys*. 16, 143–149. doi: 10.3758/BF03203267

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., and Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *J. Cogn. Neurosci*. 14, 340–347. doi: 10.1162/089892902317361886

Feigenson, L. (2007). The equality of quantity. *Trends Cogn. Sci*. 11, 185–187. doi: 10.1016/j.tics.2007.01.006

Feigenson, L., Carey, S., and Spelke, E. (2002). Infants' discrimination of number vs. continuous extent. *Cogn. Psychol*. 44, 33–66. doi: 10.1006/cogp.2001.0760

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state." A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res*. 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Fründ, I., Haenel, N. V., and Wichmann, F. A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *J. Vis*. 11, 1–19. doi: 10.1167/11.6.16

Gallistel, C. R. (1989). Animal cognition: the representation of space, time and number. *Annu. Rev. Psychol*. 40, 155–189. doi: 10.1146/annurev.ps.40.020189.001103

Gallistel, C. R. (2011). "Mental magnitudes," in *Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought*, eds S. Dehaene and L. Brannon (New York, NY: Elsevier), 3–12. doi: 10.1016/B978-0-12-385948-8.00001-3

Gallistel, C. R., and Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends Cogn. Sci*. 4, 59–65. doi: 10.1016/S1364-6613(99)01424-2

Gandini, D., Lemaire, P., and Dufau, S. (2008). Older and younger adults' strategies in approximate quantification. *Acta Psychol*. 129, 175–189. doi: 10.1016/j.actpsy.2008.05.009

Gandini, D., Lemaire, P., and Michel, B. F. (2009). Approximate quantification in young, healthy older adults, and Alzheimer patients. *Brain Cogn*. 70, 53–61. doi: 10.1016/j.bandc.2008.12.004

Geary, D. C., and Lin, J. (1998). Numerical cognition: age-related differences in the speed. *Exp. Ageing Res*. 24, 101. doi: 10.1080/036107398244274

Grondin, S., Ouellet, B., and Roussel, M.-E. (2004). Benefits and limits of explicit counting for discriminating temporal intervals. *Can. J. Exp. Psychol*. 58, 1–12. doi: 10.1037/h0087436

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., and Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proc. Natl. Acad. Sci. U.S.A*. 109, 11116–11120. doi: 10.1073/pnas.1200196109

Hasher, L., and Zacks, R. T. (1988). "Working Memory, comprehension and aging: a review and a new view," in *The Psychology of Learning and Motivation*, ed G. H. Bower (San Diego, CA: Academic Press), 193–225.

Hedden, T., and Gabrieli, J. D. E. (2004). Insights into the ageing mind: a view from cognitive neuroscience. *Nat. Rev. Neurosci*. 5, 87–96. doi: 10.1038/nrn1323

Hellström, Å. (1985). The time-order error and its relatives: mirrors of cognitive processes in comparing. *Psychol. Bull*. 97, 35–61. doi: 10.1037/0033-2909.97.1.35

Helson, H., and King, S. M. (1931). The tau effect: an example of psychological relativity. *J. Exp. Psychol*. 14, 202–217. doi: 10.1037/h0071164

Henik, A., and Tzelgov, J. (1982). Is three greater than five: the relation between physical and semantic size in comparison tasks. *Mem. Cogn*. 10, 389–395. doi: 10.3758/BF03202431

Jackson, M., and Warrington, E. K. (1986). Arithmetic skills in patients with unilateral cerebral lesions. *Cortex* 22, 611–620. doi: 10.1016/S0010-9452(86)80020-X

Javadi, A. H., and Aichelburg, C. (2012). When time and numerosity interfere: the longer the more, and the more the longer. *PLoS ONE* 7:e41496. doi: 10.1371/journal.pone.0041496

Javadi, A. H., and Aichelburg, C. (2013). Training enhances the interference of numerosity on duration judgement. *PLoS ONE* 8:e54098. doi: 10.1371/journal.pone.0054098

Karolis, V. (2013). The scale analysis of number mapping onto space: manual estimation study. *Q. J. Exp. Psychol.* 37–41. doi: 10.1080/17470218.2013.782325. [Epub ahead of print].

Kleiner, M., Brainard, D., and Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception* 36, 14. doi: 10.1016/0028-3932(83)90075-1

Lambrechts, A., Walsh, V., and van Wassenhove. (in press). Evidence accumulation in the magnitude system. *PlosOne*.

Lemaire, P., and Arnaud, L. (2008). Young and older adults' strategies in complex arithmetic. *Am. J. Psychol.* 121, 1–16. doi: 10.2307/20445440

Lemaire, P., and Lecacheur, M. (2007). Ageing and numerosity estimation. *J. Gerontol. B Psychol. Sci. Soc. Sci.* 62, P305–P312. doi: 10.1093/geronb/62.6.P305

Lourenco, S. F., Bonny, J. W., Fernandez, E. P., and Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence *Proc. Natl. Acad. Sci. U.S.A.* 109, 18737–18742. doi: 10.1073/pnas.1207212109

Lourenco, S. F., and Longo, M. R. (2010). General magnitude representation in human infants. *Psychol. Sci.* 21, 873–881. doi: 10.1177/0956797610370158

Lustig, C., and Meck, W. H. (2011). Modality differences in timing and temporal memory throughout the lifespan. *Brain Cogn.* 77, 298–303. doi: 10.1016/j.bandc.2011.07.007

Masin, S. C., and Fanton, V. (1989). An explanation for the presentation-order effect in the method of constant stimuli. *Percep. Psychophys.* 46, 483–486. doi: 10.3758/BF03210864

Meck, W. H. (2005). Neuropsychology of timing and time perception. *Brain Cogn.* 58, 1–8. doi: 10.1016/j.bandc.2004.09.004

Meck, W. H., and Church, R. M. (1983). A mode control model of counting and timing processes. *J. Exp. Psychol. Anim. Behav. Process.* 9, 320–334. doi: 10.1037/0097-7403.9.3.320

Merritt, D. J., Casasanto, D., and Brannon, E. M. (2010). Do monkeys think in metaphors? Representations of space and time in monkeys and humans. *Cognition* 117, 191–202. doi: 10.1016/j.cognition.2010.08.011

Nelson, H. E., and Willison, J. R. (1991). *National Adult Reading Test (NART).* (Windsor, UK: Nfer-Nelson).

Oliveri, M., Vicario, C. M., Salerno, S., Koch, G., Turriziani, P., Mangano, R., et al. (2008). Perceiving numbers alters time perception. *Neurosci. Lett.* 438, 308–311. doi: 10.1016/j.neulet.2008.04.051

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442. doi: 10.1163/156856897X00366

Perbal, S., Droit-volet, S., Isingrini, M., and Pouthas, V. (2002). Relationships between age-related changes in time estimation and age- related changes. *Proc. Speed Attent. Mem. Ageing Neuropsychol. Cogn.* 9, 210–216. doi: 10.1076/anec.9.3.201.9609

Pinel, P., Piazza, M., Le Bihan, D., and Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron* 41, 983–993. doi: 10.1016/S0896-6273(04)00107-2

Posner, M. I., Snyder, C. R., and Davidson, B. J. (1980). Attention and the detection of signals. *J. Exp. Psychol.* 109, 160–174. doi: 10.1037//0096-3445.109.2.160

Reynvoet, B., De Smedt, B., and Van den Bussche, E. (2009). Children's representation of symbolic magnitude: the development of the priming distance effect. *J. Exp. Child Psychol.* 103, 480–489. doi: 10.1016/j.jecp.2009.01.007

Roitman, J. D., Brannon, E. M., Andrews, J. R., and Platt, M. L. (2007). Nonverbal representation of time and number in adults. *Acta Psychol.* 124, 296–318. doi: 10.1016/j.actpsy.2006.03.008

Salthouse, T. A. (2000). Ageing and measures of processing speed. *Biol. Psychol.* 54, 35–54. doi: 10.1016/S0301-0511(00)00052-1

Salthouse, T. A., and Kersten, A. W. (1993). Decomposing adult age differences in symbol arithmetic. *Mem. Cogn.* 21, 699–710. doi: 10.3758/BF03197200

Salthouse, T. A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychol. Sci.* 2, 179–183. doi: 10.1111/j.1467-9280.1991.tb00127.x

Salthouse, T. A., Atkinson, T. M., and Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *J. Exp. Psychol. Gene.* 132, 566–594. doi: 10.1037/0096-3445.132.4.566

Sara, M., and Faubert, J. (2000). Aging, perception, and visual short-term memory for luminance-defined form. *Ophthalmic Physiol. Opt.* 20, 314–322. doi:10.1016/S0275-5408(99)00095-2

Shuman, M., and Spelke, E. (2006). Object tracking, enumeration, and individuation and element size bias numerosity perception. *J. Vis.* 6, 777. doi: 10.1167/6.6.777

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651

Tokita, M., and Ishiguchi, A. (2011). Temporal information affects the performance of numerosity discrimination: behavioral evidence for a shared system for numerosity and temporal processing. *Psychon. Bull. Rev.* 18, 550–556. doi: 10.3758/s13423-011-0072-2

Van Marle, K., and Wynn, K. (2006). Six-month-old infants use analog magnitudes to represent duration. *Dev. Sci.* 9, F41–F49. doi: 10.1111/j.1467-7687.2006.00508.x

Vanneste, S., and Pouthas, V. (1999). Timing in ageing?-the role of attention. *Exp. Aging Res.* 25, 49–68. doi: 10.1080/036107399244138

Verrillo, R. T. (1981). Absolute estimation of line length in three age groups. *J. Gerontol.* 36, 625–627. doi: 10.1093/geronj/36.5.625

Vicario, C. M., Pecoraro, P., Turriziani, P., Koch, G., Caltagirone, C., and Oliveri, M. (2008). Relativistic compression and expansion of experiential time in the left and right space. *PLoS ONE* 3:e1716. doi: 10.1371/journal.pone.0001716

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* 7, 483–488. doi: 10.1016/j.tics.2003.09.002

Wechsler, D. (1995). *Wechsler Abbreviated Scale of Intelligence. The Psychological Corporation.* New York, NY: Harcourt Brace and Company.

Xuan, B., Chen, X., He, S., and Zhang, D. (2009). Numerical magnitude modulates temporal comparison?: an ERP study. *Brain Res.* 1269, 135–142. doi: 10.1016/j.brainres.2009.03.016

Xuan, B., Zhang, D., and Chen, X. (2007). Larger stimuli are judged to last longer. *J. Vis.* 7, 1–5. doi: 10.1167/7.10.2

Xu, F., and Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition* 74, B1–B11. doi: 10.1016/S0010-0277(99)00066-9

Zorzi, M., Priftis, K., and Umiltà, C. (2002). Neglect disrupts the mental number line. *Nature* 417, 138–139. doi: 10.1038/417138a

Persistent consequences of atypical early number concepts

Michèle M. M. Mazzocco[1,2]*, Melissa M. Murphy[3], Ethan C. Brown[4], Luke Rinne[5] and Katherine H. Herold[2]

[1] Institute of Child Development, University of Minnesota, Minneapolis, MN, USA
[2] Center for Early Education and Development, University of Minnesota, Saint Paul, MN, USA
[3] School of Education, Notre Dame of Maryland University, Baltimore, MD, USA
[4] Department of Educational Psychology, University of Minnesota, Minneapolis, MN, USA
[5] School of Education, Johns Hopkins University, Baltimore, MD, USA

How does symbolic number knowledge performance help identify young children at risk for poor mathematics achievement outcomes? In research and practice, classification of mathematics learning disability (MLD, or dyscalculia) is typically based on composite scores from broad measures of mathematics achievement. These scores do predict later math achievement levels, but do not specify the nature of math difficulties likely to emerge among students at greatest risk for long-term mathematics failure. Here we report that gaps in 2nd and 3rd graders' number knowledge predict specific types of errors made on math assessments at Grade 8. Specifically, we show that early whole number misconceptions predict slower and less accurate performance, and atypical computational errors, on Grade 8 arithmetic tests. We demonstrate that basic number misconceptions can be detected by idiosyncratic responses to number knowledge items, and that when such misconceptions are evident during primary school they persist throughout the school age years, with variable manifestation throughout development. We conclude that including specific qualitative assessments of symbolic number knowledge in primary school may provide greater specificity of the types of difficulties likely to emerge among students at risk for poor mathematics outcomes.

Keywords: number concepts, place value concepts, whole number knowledge, number sense, dyscalculia, mathematics learning disabilities

INTRODUCTION

Some aspects of number knowledge involve an awareness of the meaning of somewhat arbitrary symbols (such as Arabic numerals and number words) that are used on a daily basis. This knowledge is an important predictor of later mathematics achievement (Rousselle and Noël, 2007; De Smedt et al., 2009; Krajewski and Schneider, 2009), which makes it a useful indicator of risk for mathematics learning difficulties (Mazzocco and Thompson, 2005; Jordan et al., 2007; Stock et al., 2010; Desoete et al., 2012). Although symbolic number skills begin to develop prior to schooling, they depend on instruction and typically become established in approximately first grade (Bugden and Ansari, 2011) to third grade (Girelli et al., 2000), at least for small whole numbers. Accordingly, early childhood educators' attention has been drawn to this aspect of "number sense" as a target of formal and informal learning and assessment. A challenge for educators is knowing what observable behaviors (such as counting or labeling sets) map on to important elements of number sense and how these behaviors are typically manifested in early childhood. In other words, educators may want to know what a weak number sense looks like, and which numerical behaviors reflect typical or atypical development.

One challenge in responding to this need lies in the limited delineation of number sense skills identified to date (Purpura and Lonigan, 2013), even within the subset of number skills

classified as *symbolic* representations (e.g., verbal number words, written notation, physical number lines), which collectively differ from non-symbolic arrays (e.g., visual or audible sets). Measures of number sense often represent a conglomeration of numerical tasks that vary in the degree to which they overlap with each other and with non-numerical domain-general skills such as verbal memory, working memory, or spatial reasoning (e.g., Geary, 2004). Composite standardized test scores are useful for determining broad categories of mathematics difficulties, and extreme scores may also help differentiate between children with dyscalculia—a specific mathematical learning disability—and other sources of mathematics difficulties (Rubinsten and Henik, 2009). However, the dichotomous (pass/fail) nature of the item scores used to generate composites may fail to capture meaningful differences in mathematically relevant skills between individuals at a point in time when such differentiation can aid identification and instructional priorities. Indeed, broad mathematics achievement scores may underestimate the contributions of these early foundational skills (Geary et al., 2013).

In this retrospective longitudinal study, we focus on whole number knowledge in primary school as an example of foundational skills. We focus on behaviors that would be readily assessable in informal environments, and evaluate whether early indicators of atypical number concepts are associated with future computational fluency (as proposed also by Moeller et al., 2011).

While recognizing that the number concepts we focus on are broad, we hypothesize that atypical errors on number knowledge tasks can meaningfully represent incomplete number concepts that persist throughout the school age years.

Accordingly, we propose that knowledge of such differences may be revealed through a qualitative analysis of responses to mathematics problems, with the goal of elucidating early number concepts that predict specific mathematics difficulties. We use this approach to assess aspects of performance failure rather than dichotomous pass/fail scoring, using frequency data from our completed longitudinal study to guide classification of typical and atypical errors that can then be evaluated as indicators of whole number concepts, and predictors of future math performance. The motivation for this approach is threefold: the aforementioned growing recognition that the number sense construct needs further delineation, the corresponding gaps in knowledge of developmental norms for fine grained numerical skills, and the high likelihood of behavioral differences in number skills given the heterogeneous nature of mathematical difficulties (Mazzocco, 2007; Rubinsten and Henik, 2009). We propose that the differences to emerge using this approach are likely to be meaningful indicators of later pervasive difficulties in specific areas of mathematics, because conceptual differences in number knowledge have been shown to persist well beyond the primary school age years (e.g., Mazzocco and Devlin, 2008; Geary et al., 2013).

A qualitative approach to assessing early number knowledge has both practical and theoretical significance. In practice, this approach is a complement to composite test scores for informal or formal assessments (e.g., differential diagnosis), and may be a more sensitive indicator of specific future mathematics outcomes. Theoretical contributions of qualitative error analyses provide for a more detailed understanding of developmental and individual differences in children's number sense and concepts. Although we do not claim that a qualitative approach is novel in research or assessment (e.g., Ginsburg, 2003), we do propose that it is an overlooked source of meaningful insights in the search for individual differences in number skills that do not necessarily conform to variation along a continuum. We use this approach to test the hypothesis that atypical errors that reflect numerical misconceptions in primary school are linked to aberrant responses observed late in middle school (Grade 8). This tests the broader notion that when early number concepts go awry, the consequences can persist.

MATERIALS AND METHODS
PARTICIPANTS
Participants were drawn from a longitudinal study of mathematics ability and disability described elsewhere in greater detail (Mazzocco and Myers, 2003). The initial participant pool was recruited from kindergarten classrooms in a large and socio-economically diverse public school district in the greater Baltimore, Maryland metropolitan area (which does not include schools in Baltimore city), from schools identified as having relatively low rates of mobility (to enhance retention in the longitudinal study) and low rates of free or reduced lunch participation (FRLP; as a filter for poverty). At the onset of the study, the mean FRLP rate was 16.5% (range = 1.58–29.04%) and the

mean mobility rate was also 16.5% (range = 6.8–18.9%). All 445 kindergartners with proficient English were invited to participate, and 249 (120 boys) enrolled. The sample was predominately white (86%). A total of 210 participants remained in the study for at least 4 years. The sample for the present study was drawn from this group.

The present study focused on a number writing task administered during Years 03 and 04 of the longitudinal study, when most participants were in Grades 2 and 3 (except for nine of the 210 participants who had repeated a school grade). At Grade 2 the children ranged in age from 7.0 to 8.9 years (mean = 7.78, SD = 0.34). All 210 children were included in analyses of Grade 2 and 3 math performance (eight had repeated kindergarten or Grade 1, and one repeated Grade 2).

Some of our research questions were related to mathematics achievement status, which we determined using scores from the Test of Early Mathematics Ability—Second Edition (TEMA-2) from Kindergarten to Grade 3 (described subsequently). For those analyses, 17 children met criteria for mathematics learning disability (MLD; mean age = 7.85 years), 26 met criteria for low mathematics achievement (LA; mean age = 7.92), and 123 met criteria for typical achievement in mathematics (TA; mean age 7.74 years). The remaining 44 participants were excluded from analyses based on mathematics achievement status because their TEMA-2 scores were too inconsistent over time to confidently meet criteria for MLD, LA, or TA. Thus, 166 participants were included in the final study sample for analyses pertaining to MLD status.

Finally, for analyses focused on long-term predictors of Grade 8 performance, the sample included all 153 children who participated in the overall longitudinal study through Year 09 (mean age 13.83 years). Most of these children were in Grade 8 in Year 09 of the overall study, but eight were in Grade 7 (six had repeated kindergarten or Grade 1, one repeated Grade 6, and one repeated Grade 7).

MATERIALS
Primary school mathematics tasks
Test of Early Mathematics Ability—2nd Edition. (TEMA-2; Ginsburg and Baroody, 1990). The TEMA is a standardized assessment of formal and informal mathematics knowledge normed for use with children ages 3–8 years. The TEMA-2 includes a wide range of numerical and mathematics items, such as counting aloud, counting sets, using one-to-one correspondence, number constancy, reading and writing numerals, number line concepts, and solving verbal or written arithmetic problems. Raw scores on the TEMA-2 are converted to age-referenced composites, which we used to determine participants' overall level of mathematics ability in Grades K to 3.

We used sample-based percentiles to determine mathematics ability group classification (as described elsewhere in detail, Murphy et al., 2007). Children who consistently performed below the 11th percentile on the TEMA-2 were classified as having MLD, whereas those consistently performing in the 11–25th percentile were classified as having low mathematics achievement (LA). Children consistently performing above the 25th percentile were classified as having typical achievement in mathematics (TA).

Consistency was defined as criteria being met for at least half of all years in the study, and within the 95% confidence range for all years. Note that our criteria for determining MLD status classification were aligned with reported prevalence of MLD ($\sim$6–11% as reviewed by Shalev, 2007) and we relied on sample-based vs. standard normative percentiles because our use of the TEMA-2 throughout the longitudinal study (to maintain consistency after a third edition was published) led to inflated standard scores from outdated norms.

Written Numbers Task. We focused on select number concept items given in the context of the TEMA-2 as potential predictors of later computational errors. Data came from the third and fourth years of the longitudinal study (Grades 2 and 3). For these items, children were asked to write the smallest, and the largest, one-, two-, and three-digit number, for a total of 6 trials per participant. Based on standardized scoring on the TEMA-2, there were two acceptable correct responses to the smallest one-digit number (0 and 1). The remaining five trials each had one acceptable correct response (9, 10, 99, 100, and 999, respectively). The criterion for passing the overall set of trials was six correct responses, and standardized scoring yielded one total dichotomous pass/fail score. In contrast, the scoring criteria for types of errors made were established as part of the present study, and applied individually to each trial such that the range of possible scores for number correct, number of errors, and number of specific error types was 0–6 (as described in more detail in the Results section). These scores from Grades 2 and 3 were used to predict performance on the Fast Math Test at Grade 8.

Grade 8 arithmetic fluency

Fast Math Test (FMT; Mazzocco et al., 2008). The FMT is an investigator-designed, timed, paper and pencil task used to evaluate computational fluency. In this study, we used scores from the FMT as an outcome variable in analyses with Written Number task performance as the predictor. The FMT includes 8 pages, each comprised of 18 problems, with pages alternating between two levels of difficulty (4 pages of easy problems, and 4 pages of difficult problems), two operations (4 pages of addition, 4 pages of multiplication), and two sets of identical problems presented in a different order. For each operation, "easy" problems involve one- and two-digit number combinations familiar to most middle school students that are typically solved by retrieval (e.g., 7 + 7), and "hard" problems typically require "regrouping," such that retrieval is an unlikely sole or primary strategy (e.g., 17 + 14). The FMT was administered at Grade 8 only. Test-retest reliability on this task was good. The Pearson (r) correlation between two *identical* pages was 0.83; correlations between mixed and grouped pairs of the same problem set ranged from 0.62 to 0.79 (Mazzocco et al., 2008).

Primary school and Grade 8 mathematics performance associations
Written Number and FMT performance. In this study, the outcome variable paired with the Written Numbers Task was drawn from the error coding of the FMT. In our previous work, we demonstrated that common and uncommon types of errors are observed on the FMT (Mazzocco et al., 2008). In the present study, we focused on uncommon place value errors that may represent a fundamental misconception about numbers, unlike more common miscalculation errors. Specifically, these place value errors involved numbers added across tens and ones places (referred to as NAATO errors in the original report), such as summing 6 + 2 and 1 + 3 when solving 16 + 23, thereby obtaining a sum of "48" or "84"). These errors were rarely observed among 8th graders completing the FMT, and we hypothesized that they would be related to the infrequent errors made on the Written Numbers Task—not simply because of their relative rarity but because both may reflect incomplete mastery of number concepts. Additional error types on the FMT are summarized in **Table 3**. Finally, we hypothesized that this incomplete mastery of number concepts would promote greater use of finger counting on the FMT—an infrequent strategy by 8th grade—and thus looked at the number of FMT items on which children explicitly used finger counting.

PROCEDURES
All children were tested individually by a female examiner. Parent consent and child consent were obtained in accordance to human subjects approved protocols. Testing sessions during Grades 2, 3, and 8 were approximately 90–120 min, divided into two sessions. These sessions occurred within 2 weeks of each other with rare exception; during Grade 8, some sessions occurred on the same day, pending participants' availability.

Most of the data were collected in school environments. In these cases, children were tested in their own school, in a quiet room occupied by only the student and examiner. The exception to this arrangement occurred when children moved to a non-participating school in a district for which we did not have in-school research testing privileges, or if a parent preferred to have the child tested in our laboratory. In these instances, testing occurred in a small quiet room occupied by only the student an examiner. Out-of-school testing occurred very infrequently in the primary grades, so the sample was too small to warrant statistical comparison. Upon request, Grade 8 assessments were conducted in a community-based environment (e.g., library meeting room) occupied by only the student and examiner. All data were scored, double scored, and entered twice independently and verified until all errors had been detected and corrected. Analyses were conducted using SPSS version 20 and R version 2.15.2.

RESULTS
ANALYSES RELATED TO WHOLE NUMBER KNOWLEDGE
Developmental trends and effects of mathematics ability group on total score
We ran preliminary analyses to verify anticipated effects of grade and math achievement group (TA, LA, or MLD) on overall accuracy, using a 3 (Group) $\times$ 2 (Grade) repeated measures ANOVA on the total number of correct responses (range = 0–6). Main effects were confirmed for Grade, $F_{(1, 163)} = 60.90$, $p < 0.0001$, $\eta^2 = 0.272$, with overall accuracy increasing from Grades 2 to 3 (**Table 1**); and for math achievement group, $F_{(2, 163)} = 69.08$, $p < 0.0001$, $\eta^2 = 0.418$. Children with MLD made fewer correct responses relative to the LA or TA groups, even at Grade 3,

Table 1 | Mean (and SD) number of correct responses out of 6 on Written Numbers Task among children with TA, LA, or MLD.

Participant group	Grade 2	Grade 3
TA ($n = 123$)	5.29 (0.84)	5.76 (0.67)
LA ($n = 26$)	4.81 (1.17)	5.50 (0.71)
MLD ($n = 17$)	2.59 (1.80)	4.18 (1.78)
All groups ($N = 166$)	4.93 (1.30)	5.56 (0.97)

$ps < 0.009$. The small but significant Grade × Group interaction, $F_{(2, 163)} = 7.96$, $p = 0.001$, $\eta^2 = 0.089$, reflected larger increases in accuracy over time for the MLD group, likely due to ceiling effects.

These analyses of *how many* errors children made on the Written Numbers Task reveal normal developmental trends in accuracy and quantitative differences in mathematics performance across mathematics achievement groups. **Figure 1** further illustrates developmental trends across easier-to-harder items (that is, one-, two-, and three-digit numbers) and the exaggeration of this effect in children with MLD. Whereas the effect of *grade* appears driven primarily by gains in knowledge of the largest three-digit number achieved between Grades 2 to 3, the main effect of *group* appears largely driven by the much larger proportion of children with MLD who do not make this shift at this time period.

Are these qualitative group differences significant? In both grades, many children with MLD failed even the 2-digit item(s), whereas most children with LA or TA did not. Very few children failed to identify the smallest one-digit number as either 1 or 0, but of the five that did fail, 4 had MLD (vs. 1 of 126 with TA, Fisher's Exact $p < 0.01$; and 0 of 26 with LA, Fisher's Exact $p = 0.055$). This pattern of performance veered from the more typical developmental pattern revealed by the data, and justified the following qualitative analyses.

Qualitative analyses of written numbers task errors

Does the *type* of errors made vary across math achievement group? To address this question, it was necessary to classify error types. Our *a priori* hypotheses focused on developmentally appropriate vs. idiosyncratic responses, which we operationalized in terms of *frequency* of errors made in Grade 2 across all trials and all students (210 in the study, plus 14 second graders excluded from the study due to missing data in Grade 3). Of the 1344 individual responses generated by these 224 second graders, 1068 were correct and 276 were errors. (Criteria for correct responses appear in the Methods section and in **Table 2**).

Errors were categorized as frequent or infrequent. A *frequent* error was produced by more than 3% (≥ 7) of all second graders in the study. There were four errors classified as frequent, which collectively occurred 87 times across trials and were made by 82 children. The mean number of children making any of the four frequent errors was 21.75 (range = 9–46). Thus, by definition, each frequent error was made by several children.

An error was categorized as *infrequent* if it was produced by fewer than 3% (<7) of all second graders in the study. Across all trials, 111 unique errors were classified as infrequent, which
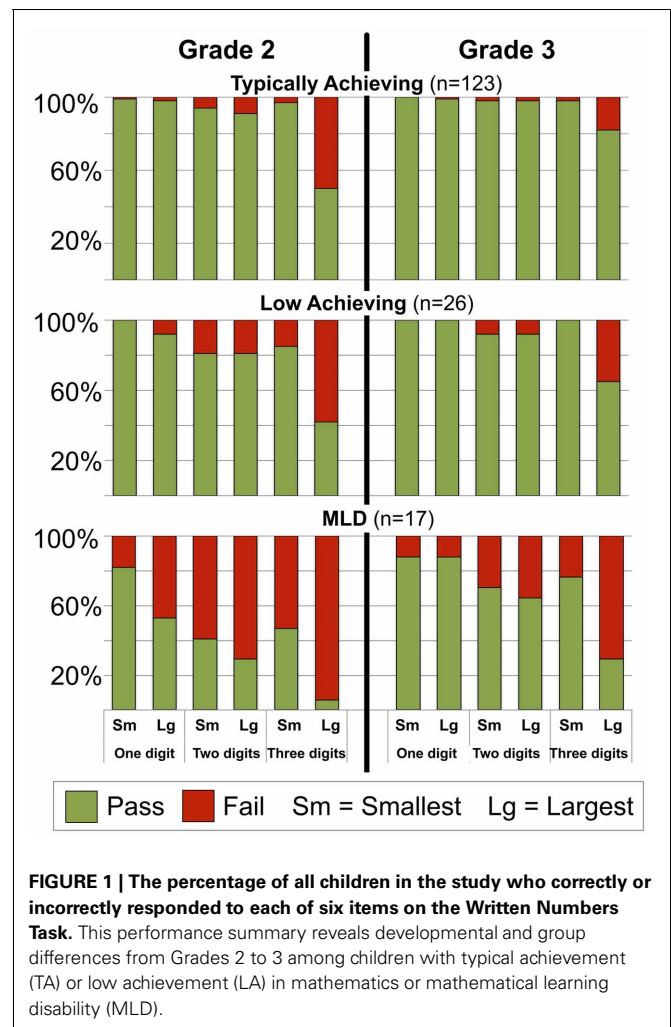
FIGURE 1 | The percentage of all children in the study who correctly or incorrectly responded to each of six items on the Written Numbers Task. This performance summary reveals developmental and group differences from Grades 2 to 3 among children with typical achievement (TA) or low achievement (LA) in mathematics or mathematical learning disability (MLD).

collectively occurred 179 times and were made by 81 children. The mean number of children making one of the 111 specific infrequent errors was 1.61, (range = 1–6). On each trial, the mean number of children who made any given infrequent error ranged from 1.32 to 2.75. Thus, by definition, infrequent errors were quite idiosyncratic in that each was made by very few children. **Table 2** includes a summary of responses observed among all 224 second graders enrolled in the larger longitudinal study.

Very few responses (10 of 1334) were reports of "I don't know" ($n = 8$) or no response at all ($n = 2$), collectively made by five children (four of whom also made infrequent errors). "I don't know" responses were neither frequent nor idiosyncratic, so they were omitted from comparisons of frequent vs. infrequent responses, which led to the exclusion of data from one child whose "I don't know" error was his only error.

Do children with MLD make significantly more infrequent responses?

We evaluated the number of infrequent errors made using a 2 (Grade) × 3 (Mathematics Achievement Group) ANOVA with only the 108 children who met criteria for MLD, LA, or TA, were tested in both Grades 2 and 3, and made at least one error in

Table 2 | Classification (and counts) of 224 second graders' responses showing correct responses, frequent errors, and a sample of infrequent errors* on individual Written Numbers Task items.

Response type	One digit				Two digit				Three digit			
	Smallest		Largest		Smallest		Largest		Smallest		Largest	
Correct	0	(118)	9	(200)	10	(192)	99	(180)	100	(196)	999	(87)
	1	(95)										
All frequent errors		(0)		(0)	11	(9)	90	(12)		(0)	199	(20)
											900	(46)
Infrequent errors												
Representative infrequent errors	2	(4)	2	(2)	19	(1)	23	(1)	102	(1)	102	(1)
	4	(1)	3	(2)	22	(2)	58	(2)	104	(1)	236	(1)
	6	(1)	12	(1)	51	(1)	89	(1)	109	(2)	308	(1)
									110	(2)	590	(1)
									197	(1)	653	(1)
									532	(1)	800	(1)
									910	(1)	901	(3)
Total infrequent errors made:		(11)		(22)		(23)		(31)		(25)		(67)
"Don't know"		(0)		(2)				(1)		(2)		(3)
No response										(1)		(1)

*By definition, frequent errors were those made by >3% of all participants (≥7/224 participants), and infrequent errors were those made by fewer than 3% (<7) of all participants. Responses of "I don't know" and non-responses were considered errors but were not classified as either frequent or infrequent.

either grade (excluding the child whose only error was an "I don't know" response). There were main effects of Grade, $F_{(1, 105)} = 34.68$, $p < 0.0001$, $\eta^2 = 0.248$, and Group, $F_{(2, 105)} = 32.55$, $p < 0.0001$, $\eta^2 = 0.383$. The number of infrequent errors decreased from Grades 2 to 3 (from 1.45 to 0.62); across grades, children with MLD made more infrequent errors (2.25) than did children with either LA (0.52) or TA (0.32), $ps < 0.0001$ (the latter two frequencies did not differ from each other, $p = 0.355$). There was a small but significant Grade × Group interaction, $F_{(2, 105)} = 2.78$, $p < 0.03$, $\eta^2 = 0.068$, reflected in **Figure 2**. Although the proportion of children making an infrequent error decreased from Grades 2 to 3, most children with MLD still did so in Grade 3. In fact, the number of infrequent errors made in Grade 3 was significantly different from zero for the MLD group only, $t_{(15)} = 3.50$, $p < 0.01$.

Perhaps the higher incidence of infrequent errors in the MLD group merely reflects greater errors of any kind among this group. If so, then their *frequent* errors should also be more prevalent. We repeated the previous 2 × 3 ANOVA, this time with the number of frequent errors as the outcome variable. Neither main effect emerged as statistically significant ($ps > 0.15$). Although the MLD group had the fewest frequent errors of any group (TA = 0.49; LA = 0.52; MLD = 0.31), these differences were not significant ($ps > 0.3$). Thus, we conclude that the performance of the MLD group is characterized by infrequent errors rather than simply more errors.

The validity of infrequent errors as indicators of atypical number concepts

Infrequent errors were most common among children with MLD (vs. LA or TA), but some children in each group made infrequent
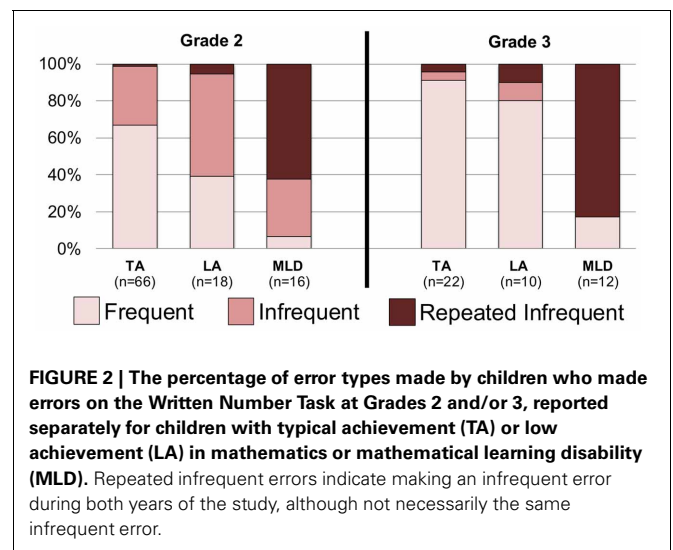
FIGURE 2 | The percentage of error types made by children who made errors on the Written Number Task at Grades 2 and/or 3, reported separately for children with typical achievement (TA) or low achievement (LA) in mathematics or mathematical learning disability (MLD). Repeated infrequent errors indicate making an infrequent error during both years of the study, although not necessarily the same infrequent error.

errors, and many of these children repeatedly made infrequent errors in Grades 2 and 3. Our subsequent analyses focused on whether children made these errors regardless of mathematics ability group.

Do infrequent errors merely reflect wild guesses? We examined metacognitive evaluation measures included in the longitudinal study test battery (Garrett et al., 2006) in which children were prompted to report if they were "sure" or "not sure" of their individual response. These prompts were administered after each trial of the Written Numbers Task. Of interest was whether children who made infrequent errors were more likely to indicate

uncertainty (i.e., to state that they were "not sure" of their response), relative to children who made frequent errors, and whether this difference was limited to instances of infrequent errors.

Using data for the trial on which infrequent responses were most common (Trial 6, the "largest three-digit number"), we carried out two contingency table analyses, one per grade. We found that nearly half (48%) of the participants who made infrequent errors were confident in their incorrect responses in Grade 2 (as indicated by their report of being "sure."), and that this rate rose to 60% among the smaller cohort that made infrequent errors in Grade 3. Although the rates were lower than the rate among respondents making frequent errors in Grade 2 (48 vs. 73%; Fisher's exact $p = 0.016$) this was not the case at Grade 3 (60 vs. 70%; $p = 0.166$), perhaps due to sample size. Nevertheless, the findings demonstrate that infrequent errors were not always merely guesses, especially in third grade, when children making these errors were more likely to be certain of their response than uncertain.

THE LONG-TERM SIGNIFICANCE OF ERROR TYPE ON THE WRITTEN NUMBERS TASK

The idiosyncratic nature of infrequent errors is perplexing, but is it meaningful? Our final set of analyses focused on whether infrequent errors in primary school are associated with computational performance at the end of Grade 8 as measured by the Fast Math Test (FMT).

First, we hypothesized that children making an infrequent error on the Written Numbers Task in Grade 3 would be more likely to make an infrequent type of place-value error on the FMT (compared to children who made either no errors or only frequent errors at Grade 3), and that this pattern would not emerge for computational errors commonly made by 8th graders. This hypothesis was supported. We found that 35% of children who made an infrequent error at Grade 3 made atypical place value errors in Grade 8, which was significantly higher than the rate of 3% among those who had not made an infrequent error in Grade 3 (odds ratio = 17.26; $p < 0.001$). Whereas 35% of children who made an infrequent error on the Grade 3 task also made

a common calculation error on the FMT (tens place addition error, defined in **Table 3**) in Grade 8, this rate did not differ from the rate of 42% among children who had not made an infrequent error at Grade 3 (odds ratio = 0.76; $p > 0.7$). The specificity is illustrated further in **Table 4**, where we also report the mean number of FMT errors and t-test results between children who did make infrequent errors at Grade 3 and those who did not.

Second, we had hypothesized that incomplete number concepts may promote finger counting during addition, and examined the number of items on which children used this strategy on the FMT. Children who committed an infrequent error in Grade 2 were more likely to use finger counting on addition problems in Grade 8 (odds ratio = 2.81; $p = 0.023$) than were those who did not make infrequent errors at Grade 2. The odds ratio based on Grade 3 data was in the same direction, but was not statistically significant (odds ratio = 2.01; $p = 0.211$, perhaps due to small sample size). These findings do not indicate causal pathways, but they do support the notion that early number concept errors have long-term consequences.

In summary, the results on the FMT analyses indicate that, relative to children who do not make infrequent errors on the Written Numbers Task, children who make infrequent errors on this task in Grades 2 or 3 are not only slower on mathematics computations in Grade 8, but also make more errors on addition and multiplication computations in Grade 8, and this higher rate of error appears to be selective.

DISCUSSION

In this retrospective study based on secondary analyses, we focused on qualitative aspects of children's early number concepts. The motivation for this study stemmed, in part, from our observations over time of the persistent difficulties some children in our longitudinal study displayed, on relatively basic arithmetic skills, from kindergarten through Grade 8. The findings show how early (and easily assessed) indicators of number skills predict later performance, thereby validating those skills as potential screening points. These findings also illustrate the value of evaluating MLD based upon specific skills rather than composite performance (e.g., Butterworth, 2005), and the implications

Table 3 | Types of errors coded on the Fast Math Test (FMT).

Operation	Error type	Definition	Example
Addition	±1	Sum is off by one (miscalculated ones place)	$7 + 12 = 18$
	±10	Sum is off by an order of ten (miscalculated tens place)	$20 + 25 = 55$
	Place value error	Adding numbers across tens and ones places (in the example, $7 + 2 = 9$, and $1 + 3 = 4$)	$17 + 23 = 94$ $17 + 23 = 49$
Multiplication	Operand	A problem is solved using the wrong operand	$5 \times 7 = 30$
	Table	The given answer is a product on the 10×10 multiplication table, but not the correct product	$8 \times 8 = 63$
	Non-table	The given answer is not a product on the 10×10 multiplication table	$7 \times 3 = 26$
	Regrouping	Incomplete or incorrect regrouping	$10 \times 8 = 108$
Either operation	Operation	A problem is solved using the wrong operation	$4 + 3 = 12$ $4 \times 3 = 7$

Adapted from a table appearing in Mazzocco et al. (2008).

Table 4 | Mean (and SD) number of errors made on 8th grade timed calculation on the Fast Math Test (FMT), as a function of types of calculation errors and whether infrequent errors were made on the Written Numbers Task at Grade 2 or Grade 3.

	Infrequent errors made in Grade 2?			Infrequent errors made in Grade 3?		
	No (n = 102)	Yes (n = 51)	p	No (n = 136)	Yes (n = 17)	p
MULTIPLICATION						
Operand	1.26 (1.49)	1.92 (2.12)	0.052[†]	**1.35 (1.63)**	**2.59 (2.29)**	**0.005**
Table	0.55 (0.97)	0.86 (1.40)	0.108	0.58 (1.04)	1.24 (1.68)	0.134[†]
Non-Table	0.79 (1.58)	0.86 (1.43)	0.794	0.76 (1.55)	1.29 (1.31)	0.173
Regrouping	0.61 (0.99)	0.51 (0.73)	0.531	0.54 (0.92)	0.82 (0.81)	0.233
Total	3.40 (3.24)	4.51 (4.32)	0.110[†]	3.41 (3.34)	6.65 (4.83)	0.000
ADDITION						
±1	0.79 (1.10)	1.16 (1.35)	0.077	0.86 (1.19)	1.35 (1.22)	0.110
±10	0.62 (0.95)	0.84 (1.36)	0.293[†]	0.68 (1.05)	0.82 (1.51)	0.607
Place value	**0.03 (0.17)**	**0.22 (0.61)**	**0.037[†]**	**0.03 (0.17)**	**0.59 (0.94)**	**0.026[†]**
Total	2.75 (2.25)	4.12 (4.51)	0.044[†]	2.92 (2.62)	5.47 (5.93)	0.098[†]
EITHER OPERATION						
Operation	0.39 (0.88)	0.65 (0.89)	0.095	0.40 (0.80)	1.06 (1.30)	0.058[†]
RT on easy sets (seconds)	**128.9 (31.6)**	**153.0 (43.9)**	**0.001[†]**	**131.8 (33.8)**	**177.9 (43.6)**	**0.000**

[†]Reflects Satterthwaite approximation for degrees of freedom to correct for unequal variances; used only in cases of unequal variance.

of individual differences for persistent mathematics difficulties. Specifically, we show that the occurrence and nature of atypical number concepts at Grade 3 are associated with accuracy and types of errors made on mental calculations at Grade 8. Finally, our findings support claims of qualitative differences in early number skills between children with vs. without MLD (e.g., Mazzocco et al., 2011), although claims counter to this notion have also been supported (e.g., Landerl and Kölle, 2009). We believe these findings have practical and theoretical value, despite their preliminary nature.

We examined the "smallest" and "largest" numbers children generated in second and third grade, a time by which we would expect children's single digit whole number knowledge to be well-mastered, their associations between symbolic numbers and meanings to be automatized (e.g., Girelli et al., 2000), and their familiarity with numbers to apply to 2- and 3-digit numbers. Several interesting patterns emerged from this study related to developmental and individual differences in performance accuracy on this task and to the implications of these differences for future computational fluency. First, we found anticipated developmental trends in how accurately children generated two- and three-digit numbers from Grades 2 to 3. The pattern of heightened accuracy in two- and, later, three-digit numbers parallels findings that two-digit processing is not automatically generalized to three-digit number processing (Mann et al., 2012). Next, and as anticipated, these developmental differences seen in overall accuracy were somewhat exaggerated in children with low achievement, and remarkably exaggerated among children with MLD, who continued to make errors at a much higher rate through Grade 3 (**Figure 1** depicts both the developmental trends and the group differences). More importantly, interesting individual differences were observed in the nature of children's errors when generating two- and three-digit numbers.

Not all errors on the Written Numbers Task were equivalent. Some implicated developmental trends that have not been reported previously, but which align with recent evidence of a hundreds-place focus when processing three-digit numbers, at least among children in this age group (Mann et al., 2012). For instance, the most frequent *incorrect* response to prompts for the largest three-digit number was "900," which was reported by 46 of the 224 s. graders. The number "90" was the only frequent error to prompts for the largest two-digit number (although made by only 12 s graders). Frequent errors were less common in children with MLD, despite the fact that children with MLD made more errors overall. Only one child with MLD reported that "900" was the largest three-digit number, and yet (as seen in **Figure 1**), most children with MLD made errors on this trial. Although only a small group of children repeatedly made idiosyncratic errors (that is, in both grades), this was a characteristic of most of the children with MLD (**Figure 2**). Note, however, that children rarely reported the same idiosyncratic answer in both years.

Beyond the mere emergence of these idiosyncratic errors, of particular interest is the finding that such errors observed at Grade 3 were associated with specific and unusual errors on mathematics computation 5 years later. On the one hand, these results support the notion that idiosyncratic number concepts in early childhood are a meaningful reflection of persistent number concept anomalies which may affect the foundation for later arithmetic computation (as proposed by Mann et al., 2012). On the other hand, our interpretation is far from definitive given the sample size and other limitations associated with our retrospective analysis of longitudinal data collected for a prospective study.

It is possible that the primary predictor variable that we explored here—infrequent errors in written whole numbers—is simply a repackaging of the MLD criteria used in the study. Based on these criteria, children with MLD make more infrequent errors

than their primary school peers on the Written Numbers Task, they make more place value errors than their 8th grade peers (Mazzocco et al., 2008), and they are generally less accurate at evaluating their own math performance and thus produce ratings of confidence poorly calibrated with their performance (Garrett et al., 2006). Yet infrequent errors were also made by some (albeit, very few) children with LA or TA, and a few children with MLD did not make any infrequent errors. Additional support that MLD and infrequent error criteria are distinct (even if overlapping) predictors comes from the finding that infrequent errors at Grade 3 did not predict all types of Fast Math Test errors that occur with greater frequency among children with MLD.

CONCLUSIONS AND FUTURE DIRECTIONS

In this retrospective study, we demonstrate how a qualitative error analysis of early symbolic number knowledge reveals potential sources of individual differences that may affect mathematics outcomes 6 years later. This means that misguided early number concepts may have long-term implications. Our goal was not to definitively identify core deficits of dyscalculia or MLD, but rather to illustrate the contribution of qualitative analysis to delineating meaningful aspects of number sense by focusing on one representation of number concepts.

Although we focused on qualitative analysis of errors, *correct* responses may also be revealing. For instance, most second graders (118 of 224) reported that the smallest 1 digit number was "0," and fewer than half (95) reported this value was "1." In contrast, among children with MLD who correctly answered this item, most responded "1" rather than "0." Both answers are scored as correct, but each may represent different levels of understanding of these small numbers. At issue is how well responses such as these reflect the nature of children's fundamental number concepts.

Such qualitative analyses of responses, including errors, must be evaluated relative to developmental norms. For instance, over time, errors that were considered frequent vs. infrequent must be re-evaluated. In our longitudinal study, we continued to administer the Written Numbers Task to children who failed any of the six trials in a given year, so we were able to discover that of the seven children who continued to err on trial 6 during Grade 5, all answered "900." [Some children continued to err on this item

in Grade 6 ($n = 6$), Grade 7 ($n = 3$), and even Grade 8 ($n = 1$)]. Whereas at Grade 3 this response was categorized as a *frequent* error, it became *infrequent* by Grade 5.

Our observations underscore recommendations for the use of thoughtful questioning in mathematics assessments and teaching (Ginsburg, 2003) and when seeking to differentiate delayed vs. deficient mathematics skills (Rubinsten and Henik, 2009), especially when compensatory mechanisms mask otherwise aberrant numerical processing (Murphy and Mazzocco, 2008). Teachers can remain attentive for atypical errors on an informal basis as a source of information used to guide their online and systematic decision-making about students' individual learning needs or difficulties. Information about *where* trouble may occur down the line increases the specificity of targeted interventions. Since early basic number knowledge deficits can persist throughout the school age years, we must be mindful that their manifestation varies with development, and that the inclusion of specific qualitative assessments of symbolic number knowledge in primary school can provide greater specificity of the types of difficulties likely to emerge among students at risk for poor mathematics outcomes.

AUTHOR CONTRIBUTIONS

Michèle M. M. Mazzocco conceived and designed the study; Michèle M. M. Mazzocco and Melissa M. Murphy carried out the study; Michèle M. M. Mazzocco and Ethan C. Brown analyzed data; Michèle M. M. Mazzocco, Melissa M. Murphy, and Ethan C. Brown wrote the paper. All authors contributed to editing and reviewing the paper.

ACKNOWLEDGMENTS

REFERENCES

Bugden, S., and Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition* 118, 32–44. doi: 10.1016/j.cognition.2010.09.005

Butterworth, B. (2005). "Developmental dyscalculia," in *Handbook of Mathematical Cognition*, ed J. I. D. Campbell (New York; Hove; East Sussex: psychology Press), 455–467.

De Smedt, B., Verschaffel, L., and Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *J. Exp. Child Psychol.* 103, 469–479. doi: 10.1016/j.jecp.2009.01.010

Desoete, A., Ceulemans, A., De Weerdt, F., and Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten. Findings from a longitudinal study. *Br.*

J. Educ. Psychol. 82, 64–81. doi: 10.1348/2044-8279.002002

Garrett, A. J., Mazzocco, M. M. M., and Baker, L. (2006). Development of the metacognitive skills of prediction and evaluation in children with or without math disability. *Learn. Disabil. Res. Pract.* 21, 77–88. doi: 10.1111/j.1540-5826.2006.00208.x

Geary, D. C. (2004). Mathematics and learning disabilities. *J. Learn. Disabil.* 37, 4–15. doi: 10.1177/00222194040370010201

Geary, D. C., Hoard, M. K., Nugent, L., and Bailey, D. H. (2013).

Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS ONE* 8:e54651. doi: 10.1371/journal.pone.0054651

Ginsburg, H. P. (2003). *Test of Early Mathematics Ability Third Edition Assessment Probes and Instructional Activities.* Austin, TX: Pro-Ed.

Ginsburg, H. P., and Baroody, A. J. (1990). *Test of Early Mathematics Ability, 2nd Edn.* Austin, TX: Pro-Ed.

Girelli, L., Lucangeli, D., and Butterworth, B. (2000). The

development of automaticity in accessing number magnitude. *J. Exp. Child Psychol.* 76, 104–122. doi: 10.1006/jecp.2000.2564

Jordan, N. C., Kaplan, D., Locuniak, M. N., and Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learn. Disabil. Res. Pract.* 22, 36–46. doi: 10.1111/j.1540-5826.2007.00229.x

Krajewski, K., and Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: Findings from a four-year longitudinal study. *Learn. Instr.* 9, 513–526. doi: 10.1016/j.learninstruc.2008.10.002

Landerl, K., and Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. *J. Exp. Child Psychol.* 103, 546–565. doi: 10.1016/j.jecp.2008.12.006

Mann, A., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, C. (2012). On the development of Arabic three digit number processing in primary school children. *J. Exp. Child Psychol.* 113, 594–601. doi: 10.1016/j.jecp.2012.08.002

Mazzocco, M. M. M. (2007). "Defining and differentiating mathematical learning disabilities and difficulties," in *Why is Math so Hard for Some Children? The Nature and Origins of Mathematics Learning Difficulties and Disabilities,* eds D. B. Berch and M. M. M. Mazzocco (Baltimore, MD: Paul H. Brookes), 29–47.

Mazzocco, M. M. M., and Devlin, K. T. (2008). Parts and 'holes': Gaps in rational number sense among children with vs. without mathematical learning disabilities. *Dev. Sci.* 11, 681–691. doi: 10.1111/j.1467-7687.2008.00717.x

Mazzocco, M. M. M., Devlin, K. T., and McKenney, J. L. (2008). Is it a fact. Timed arithmetic performance of children with mathematical learning disabilities (MLD) varies as a function of how MLD is defined. *Dev. Neuropsychol.* 33, 318–344. doi: 10.1080/87565640801982403

Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability. *Child Dev.* 82, 1224–1237. doi: 10.1111/j.1467-8624.2011.01608.x

Mazzocco, M. M. M., and Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school age years. *Ann. Dyslexia* 53, 218–253. doi: 10.1007/s11881-003-0011-7

Mazzocco, M. M. M., and Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learn. Disabil. Res. Pract.* 20, 142–155. doi: 10.1111/j.1540-5826.2005.00129.x

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., and Nuerk, H.-C. (2011). Early place-value understanding as a precursor for later arithmetic performance—A longitudinal study on numerical development. *Res. Dev. Disabil.* 32, 1837–1851. doi: 10.1016/j.ridd.2011.03.012

Murphy, M. M., and Mazzocco, M. M. (2008). Rote numeric skills may mask underlying mathematical disabilities in girls with fragile X syndrome. *Dev. Neuropsychol.* 33, 345–364. doi: 10.1080/87565640801982429

Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., and Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *J. Learn. Disabil.* 40, 458–478. doi: 10.1177/00222194070400050901

Purpura, D. J., and Lonigan, C. J. (2013). Informal numeracy skills: the structure and relations among numbering, relations, and arithmetic operations in preschool. *Am. Edu. Res. J.* 50, 178–209. doi: 10.3102/0002831212465332

Rousselle, L., and Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non symbolic number magnitude processing. *Cognition* 102, 361–395. doi: 10.1016/j.cognition.2006.01.005

Rubinsten, O., and Henik, A. (2009). Developmental dyscalculia: heterogeneity might not mean different mechanisms. *Trends Cogn. Sci.* 13, 92–99. doi: 10.1016/j.tics.2008.11.002

Shalev, R. S. (2007). "Prevalence of developmental dyscalculia," in *Why is Math So Hard for Some Children? The Nature and Origins of Mathematical Learning Difficulties and Disabilities,* eds D. B. Berch and M. M. Mazzocco (Baltimore, MD: Paul H Brookes Publishing), 49–60

Stock, P., Desoete, A., and Roeyers, H. (2010). Detecting children with arithmetic disabilities from kindergarten: evidence from a 3-year longitudinal study on the role of preparatory arithmetic abilities. *J. Learn. Disabil.* 43, 250–268. doi: 10.1177/0022219409345011

Quantitative deficits of preschool children at risk for mathematical learning disability

Felicia W. Chu[1], Kristy vanMarle[1] and David C. Geary[1,2]*

[1] Department of Psychological Sciences, University of Missouri, Columbia, MO, USA
[2] Interdisciplinary Neuroscience Program, University of Missouri, Columbia, MO, USA

The study tested the hypothesis that acuity of the potentially inherent approximate number system (ANS) contributes to risk of mathematical learning disability (MLD). Sixty-eight (35 boys) preschoolers at risk for school failure were assessed on a battery of quantitative tasks, and on intelligence, executive control, preliteracy skills, and parental education. Mathematics achievement scores at the end of 1 year of preschool indicated that 34 of these children were at high risk for MLD. Relative to the 34 typically achieving children, the at risk children were less accurate on the ANS task, and a one standard deviation deficit on this task resulted in a 2.4-fold increase in the odds of MLD status. The at risk children also had a poor understanding of ordinal relations, and had slower learning of Arabic numerals, number words, and their cardinal values. Poor performance on these tasks resulted in 3.6- to 4.5-fold increases in the odds of MLD status. The results provide some support for the ANS hypothesis but also suggest these deficits are not the primary source of poor mathematics learning.

Keywords: approximate number system, quantitative knowledge, mathematics achievement, learning disability, dyscalculia, Title I preschool, executive control

INTRODUCTION

The poor mathematical skills of nearly one in four adults in many modern economies places them at heightened risk for underemployment and frequent unemployment, controlling for reading ability, intelligence, and race (Rivera-Batiz, 1992; Bynner, 1997). These employment-related risks can, in many cases, be traced to poor numerical competencies at school entry (Duncan et al., 2007; Morgan et al., 2009); children who start school with a poor understanding of numerals are four times more likely than their peers to score in the bottom quartile on employment-relevant quantitative tests by adolescence, controlling for other factors (Geary et al., 2013). Consequently, reducing this long-term risk may require identification and eventually the remediation of prekindergarten precursors of these school-entry deficits.

One hypothesis is that poor school-entry number knowledge results from deficits in systems for representing the approximate quantity of collections of items [approximate number system (ANS); Gilmore et al., 2010; Piazza et al., 2010; Mazzocco et al., 2011b] or the ability to quickly apprehend the exact quantity of small sets of items (subitizing; Koontz and Berch, 1996; Landerl et al., 2003; Butterworth, 2005; Iuculano et al., 2008). It is currently debated whether subitizing is dependent on the ANS or is a distinct system, and for ease of presentation we assume a single system. The ANS is thought to underlie humans' intuitive sense of numerical magnitude (Gallistel and Gelman, 1992, 2000; Feigenson et al., 2004). We tested the hypothesis that a deficit in this system contributes to poor mathematics achievement by comparing preschool children at risk for a mathematical learning disability (MLD) to a group of their typically achieving (TA) peers on a measure of ANS acuity. As a contrast to any ANS deficit, the groups

were also compared on other quantitative competencies that have been shown to be predictive of later mathematics achievement.

The acuity of the ANS is assessed by children's sensitivity to subtle differences in the relative magnitudes of collections of objects, and individual differences in this sensitivity may be correlated with mathematics achievement. For example, ninth graders' ANS acuity was found to be retrospectively correlated with standardized mathematics achievement scores as far back as kindergarten, controlling for intelligence, executive functions, and other factors (Halberda et al., 2008). A follow-up study revealed that ANS acuity was particularly poor for adolescents with MLD, again controlling for working memory and other factors (Mazzocco et al., 2011a). In addition, some studies have found that school-age children with MLD may have deficits or delays in the ANS system (Landerl et al., 2003; Mazzocco et al., 2011b). Piazza et al. (2010) found that ANS sensitivity of 10-year-olds with MLD was about the same as that of TA 5-year-olds matched on intelligence.

Other studies, however, have found no evidence of an ANS deficit in children with poor mathematics achievement generally or MLD in particular (Rousselle and Noël, 2007; Iuculano et al., 2008). These studies and related ones suggest that individual differences in mathematics achievement and MLD may instead be due to deficits or delays in the explicit attentional sensitivity to or understanding of Arabic numerals, number words, and the relations among them (Hannula et al., 2010; Bugden and Ansari, 2011; De Smedt and Gilmore, 2011; Lyons and Beilock, 2011). The extent to which ANS sensitivity contributes to the initial learning of this explicit quantitative knowledge remains to be determined (Gilmore et al., 2010; Geary, 2013), but at the very least these studies indicate that the testing of the ANS hypothesis needs to be

done in the context of a broad assessment of basic quantitative knowledge and skills. In this way, the relative contribution of ANS sensitivity to poor mathematics achievement can be contrasted with the relative contribution of other quantitative competencies.

Candidates for these other competencies include those that have been identified in studies of infants and preschool children (e.g., Gelman and Gallistel, 1978; Strauss and Curtis, 1984; Wynn, 1990; vanMarle, 2013), especially those with a demonstrated link to later mathematics achievement (Locuniak and Jordan, 2008; Jordan et al., 2009b; Krajewski and Schneider, 2009; LeFevre et al., 2010). For instance, using quantitative tasks that assess children's skills at counting objects, knowledge of counting principles, numeral recognition, and simple non-verbal addition and subtraction, Jordan and colleagues found that early mathematical knowledge and growth in this knowledge was predictive of mathematics achievement in second and third grade (Locuniak and Jordan, 2008; Jordan et al., 2009b).

The preschool predictors of risk for later MLD are not currently known, but children scoring in the bottom 25% on mathematics achievement tests and especially those scoring in the bottom 10% are likely at risk (Geary et al., 2007; Murphy et al., 2007). These cutoffs are based on studies of school age children and adolescents who have difficulties learning mathematics. Students with MLD include as many as 7% of children and adolescents (ranging from 4 to 14% depending on classification methods), and consistently (across grades) score at or below the 10th-percentile on mathematics achievement tests (Lewis et al., 1994; Barbaresi et al., 2005; Shalev et al., 2005). An additional 10% or so of children are persistently low achieving (LA) and score between the 11th and 25th percentiles in mathematics across grades, despite average intelligence and reading ability (for reviews, see Dowker, 2005a; Berch and Mazzocco, 2007).

In a 5-year prospective study, Geary et al. (2012b) found that children with MLD and their LA peers represented different cut points on the normal distribution of mathematical achievement. For some numerical or arithmetical competencies the children in these groups showed similar deficits, relative to TA children, whereas in others the LA children showed more rapid across-grade gains than the children with MLD. Other studies have also found different numerical and arithmetical patterns of strengths and deficits within MLD and LA groups, and even within TA groups (Denvir and Brown, 1986; Dowker, 2005b; Jordan et al., 2009a; Geary et al., 2012a). The results suggest that MLD and LA may represent different levels of severity for some basic numerical deficits, different patterns of deficit for others, and that even within such groups there are often quantitative strengths as well as deficits.

The present study focused on the quantitative development of children at the beginning and the end of their first year of Title I preschool. Title I is a federally (United States) funded program that provides services to children at risk for school failure, and thus includes a disproportionate number of children who are likely at risk for later MLD or LA (hereafter, MLD). In addition to a broad assessment of quantitative competencies, including ANS sensitivity, we also assessed other factors that have been shown to influence mathematical learning; specifically, intelligence (Deary et al., 2007; Geary, 2011) and executive functioning (Blair and Razza, 2007; Bull et al., 2008; Clark et al., 2010). These were used as covariates

in our contrasts of MLD and TA groups, along with measures of preliteracy skills and parental education.

MATERIALS AND METHODS
PARTICIPANTS
Seventy-one children were recruited from the Title I preschool program within the public school system in Columbia, MO, USA; data for two children were excluded due to very low (<61) intelligence scores and one other moved. The results presented are based on the remaining 68 (35 boys) children. Title I Preschool is a federally funded program offering services to 3- to 5-year-old children with developmental needs, and is designed to prepare them for successful school entry. The Columbia Public Schools Title I Preschool program serves about 750 children, with 26 classrooms located throughout the district. Consent forms were sent to all entering 3-year-olds (~240 children), and the sample consisted of those whose parents consented to participation. At the time of the first assessment, the children were 3 years 9 months of age (range: 3y2m–4y2m).

Demographic information was obtained through parent survey for a subset ($n = 51$) of the sample. Of those parents who returned the survey, not all provided responses to all questions and thus the number of responses varied by question. The ethnic composition of the sample was 81% non-Hispanic, 10% Hispanic/Latino, and the remaining unknown. The racial composition was 60% White, 20% Black, 8% Asian, 8% more than one race, and 4% unknown. The self-reported total household income was: $0–$25k (35%), $25k–$50k (21%), $50k–$75k (25%), $75k–100k (15%), $100k–$150k (2%), $150k or more (2%). Thirty-three percent of respondents reported receiving food stamps, and 8% reported receiving housing assistance.

From the survey, we were most interested in parental education ($n = 49$). The highest level of mothers' education was: some high school (10%), complete HS/GED (48%), bachelor's degree (30%), post-graduate degree (12%). The highest level of fathers' education was some high school (12%), complete HS/GED (43%), bachelor's degree (18%), post-graduate degree (27%). Maternal and paternal education levels were found to be highly correlated, $r_{48} = 0.80$, $p < 0.0001$, and thus we created a mean parental education variable ($\alpha = 0.88$). Three groups were then created from this variable: no information provided ($n = 19$), high school ($n = 25$), and college ($n = 24$). The high school group consisted of children who had at least one parent with a high school diploma or equivalent but no better, and the college group consisted of those with at least one parent who was a college graduate (or better).

MATERIALS
Quantitative tasks
Our quantitative tasks were administered in two sessions, each conducted once in Fall and once in Spring, for a total of four sessions. These sessions assessed verbal and non-verbal counting, numeral recognition, ordinality, cardinality, magnitude sensitivity, and informal arithmetic.

Counting. We assessed children's conceptual knowledge and procedural skills using three tasks: enumeration, verbal counting, and counting knowledge. For the *enumeration* task, children were

shown an array of 20 stickers and asked to count them, pointing to each one. The score was the highest number counted before committing an error. The *verbal counting* task involved the child reciting the count list, starting from "one" and counting as high as they could without an error, or until they reached 100. This task determined how well the child had the count list committed to memory. The *counting knowledge* task assessed children's understanding of basic counting concepts (e.g., one-one correspondence; Gelman and Gallistel, 1978) and their awareness of essential and unessential features of counting (Briars and Siegler, 1984). On each of 13 trials, children watched a puppet count a line of checker pieces (alternating in color, red and black) and then indicated whether the count was "OK," or "Not OK and wrong." There were four types of trials: correct (four trials), right-left (four trials), pseudo-error (five trials), and error (four trials) (Geary et al., 1992). For correct trials, checkers were counted sequentially and correctly, from left to right. In right-left trials, checkers were counted sequentially and correctly, from right to left. Pseudo-error trials consisted of counting the pieces correctly from left to right, starting first with one color and then returning to the left side of the array and continuing with the other color. For error trials, checkers were counted sequentially from left to right, but the first checker was counted twice. The score was the overall percent of trials correctly identified as "OK" (i.e., correct, right-left, pseudo-error) or "not OK and wrong" (i.e., error).

Numeral recognition. For the *numeral recognition* task, children were shown the Arabic numerals (one-at-a-time) from 1 to 15 in random order. Children were asked to name each one, and the score was the total number of numerals correctly named. Only the numerals correctly identified were used in the *numeral comparison* task (below).

Ordinality. Two tasks were used to assess ordinality. The *numeral comparison* task targeted children's understanding of ordinality by asking them to compare two Arabic numerals and report "which is bigger?" Each child completed six comparisons. The score was the total percent correct across the six trials. This task tested whether children have mapped Arabic numerals onto non-verbal quantities and whether they understand the numerals as an ordered sequence.

The second task was the *ordinal choice* task. This task was based on a common procedure that has been used successfully with preverbal infants with small (Feigenson et al., 2002) and large set sizes (vanMarle and Wynn, 2011; vanMarle, 2013) and with non-human primates (vanMarle et al., 2006). Children watched an experimenter sequentially hide two different numbers of objects (e.g., small toy fish) in two opaque cups; items were dropped into the cups one at a time. The children were then asked to pick the cup that contained more objects. There were six different comparisons (1 vs. 2, 2 vs. 3, 3 vs. 4, 4 vs. 5, 5 vs. 6, and 6 vs. 7). In order to successfully identify the larger quantity, children had to mentally track the sum and compare the number of objects in each cup. Because the comparisons varied in difficulty (i.e., ratios varied from 0.5 to 0.86), we generated a single score that was weighted for the difficulty

of the comparison. This was done by first multiplying each trial's score (incorrect = 0, correct = 1) by the ratio of the comparison (e.g., 2 vs. 3 = 0.67) and then summing the products across trials.

Cardinality. Children completed two tasks that assessed their knowledge of cardinal value (Wynn, 1990; Sarnecka and Carey, 2008). In the *give-a-number* task (Wynn, 1990), children were asked to give the experimenter exactly 1, 2, 3, 4, 5, and 6 objects from a pile. Children began at set size 1 and advanced to the next set size after a correct response; if they were incorrect, they went down one set size. The highest number of objects they correctly gave the experimenter on at least two of three attempts was taken as the highest set size for which the child understood cardinality (Le Corre and Carey, 2007).

The second task, *point-to-x* (Wynn, 1990), required children to "point to the picture that has *x* objects." Children received two blocks of six trials with ratios ranging from 0.5 to 0.67 (1 vs. 2, 5 vs. 10, 2 vs. 3, 6 vs. 9, 4 vs. 7, and 5 vs. 8), with both exclusively large and small sets represented. On each trial, children saw two sets of pictured objects on a laptop display (one on the left and one on the right). The smaller number was the target on half of the trials, and the side on which the smaller set was displayed was counterbalanced across trials. The score was determined by multiplying each trial's score (incorrect = 0 or correct = 1) by the ratio of the comparison (e.g., 5 vs. 10 = 0.5). Products were summed across trials to produce a single score weighted for the difficulty of the comparison.

Magnitude sensitivity. Magnitude sensitivity was tested using a discrete quantity discrimination task (hereafter, ANS task) and a continuous quantity discrimination task. The ANS task assessed the precision with which children mentally represent discrete quantities of objects. Using the Panamath program (Halberda et al., 2008), children received 24 test trials on a laptop computer. Each trial contained two sets of blue and yellow dots (each set was contained within a rectangle), and children identified which set "had more dots." All dot displays consisted of more than three dots and were displayed for only 2533 ms in order to discourage verbal counting. Ratios of blue:yellow dots were randomly selected for each trial and varied between 1.29 and 3.38. The Panamath program provides estimates of children's Weber fraction (w), which is thought to index ANS acuity or the precision with which one can represent a given quantity, and percent correct (see Halberda and Feigenson, 2008).

The continuous quantity discrimination task was conceptually similar, but children were asked to discriminate a continuous quantity, surface area. Children were presented with 24 test trials; in each trial, they were presented with a rectangle made of blue and red squares (four trials at each of six red:blue ratios – 1:4, 1:3, 1:2, 2:3, 3:4, and 4:5). For each trial, children reported whether there was "more red" or "more blue" in the picture.

Informal arithmetic. Children's early arithmetic skills were assessed using two tasks. The *magic box* task (vanMarle and Wynn, 2001) is a variant of Starkey's (1992) search-box task and was designed to assess children's implicit understanding of addition

and subtraction. Children were first introduced to a puppet that they were told would sometimes perform a magic trick on items hidden in a box. Unbeknownst to the child, a false floor inside the box could be manipulated to create the illusion that an object had appeared or disappeared when the lid was closed. On each trial, the child watched an experimenter hide 0, 1, or 2 objects in the box. The experimenter then added or removed an object from the hidden set as the child watched. Children were not allowed to see the resulting set. The lid was closed and then opened to reveal either the correct result of the operation, or the incorrect result. There were eight trials in this task, with a correct and an incorrect result for each of four problems: $0 + 1 = 1$ or 0, $1 + 1 = 2$ or 1, $1 - 1 = 0$ or 1, $2 - 1 = 1$ or 2. When the result was revealed, children were asked whether the puppet had done a magic trick. In order to correctly identify the incorrect outcomes as magical and the correct outcomes as not magical, the child needed to understand the effects of addition or subtraction of an object on set size. This task only required children to detect whether the operation was correct or incorrect, and did not require them to predict the exact result of the operation.

The second task was *non-verbal calculation* (Levine et al., 1992). Here, children were shown addition or subtraction of one or more disks from a hidden set of disks and then asked to predict the exact numerical result. Children watched an experimenter place a number of plastic disks in a line; the experimenter then covered the disks with a plate and added or removed some from under the plate. Children were asked to create a set of disks equal in number to the hidden set, but could also report the answer verbally. After four familiarization trials in which the children simply matched a hidden set, there were 12 test trials, presented in random order: $3 - 1$, $2 + 2$, $4 - 2$, $1 + 3$, $4 - 1$, $4 + 1$, $3 + 2$, $1 + 4$, $5 - 2$, $5 - 3$, $2 + 4$, and $6 - 4$.

Cognitive measures

Children completed a cognitive battery to control for intelligence, executive control, and preliteracy skills.

Intelligence. The children were administered the Receptive Vocabulary, Block Design, and Information subscales of the *Wechsler Preschool and Primary Scale of Intelligence – III* (WPPSI; Wechsler, 2002). Following standard procedures, scores were scaled and prorated to generate an estimate of Full Scale IQ (intelligence).

Executive control. Executive function was assessed using the Conflict EF scale developed for children from 2 to 6 years of age (Beck et al., 2011). This scale consists of six levels; the first four included two subsections (five trials each), whereas Levels 5, 6A, and 6B included 10 trials each. All children began on Level 2 following age-based procedures.

The Conflict EF scale consisted of a card-sorting task. Children were presented with two black plastic index card boxes with holes cut into the top; each box had a target card affixed to the front. Children were given a rule and asked to place a card in the appropriate box. Each level consisted of normal sorting trials, followed by conflict trials. For example, children placed the card in the corresponding box depending on whether the card was a "big kitty" or a "little kitty"; in conflict trials, children were asked to switch

the rule, i.e., a "big kitty" would go in the "little kitty" box. In subsequent levels, children sorted the cards depending on shape or color of the card (again, the rule was reversed to create a conflict trials). More advanced levels required children to sort cards according to shape or color depending on whether a black border was present or absent on the card. In order to move on to the next level, children had to complete four out of five trials correctly; in levels with 10 trials, children had to complete four shape trials and four color trials correctly in order to proceed to the next level. The score was the total number of correct conflict trials.

Preliteracy. To assess children's preliteracy skills, one subtest of the *Phonological Awareness Literacy Screening-PreK* (PALS; Invernizzi et al., 2004), Upper-Case Alphabet Recognition, was administered. This task was chosen because it is a reliable indicator of later reading ability (Blatchford et al., 1987). Children were presented with capital letters in the alphabet (a few at a time) and asked to identify each letter. The score was the total number of letters correctly identified.

Mathematics achievement groups

In order to identify mathematics achievement groups, participants also completed the *Test of Early Mathematical Ability-3* (TEMA-3; Ginsburg and Baroody, 2003), which is a nationally normed ($M = 100$, SD = 15) measure of young children's mathematical competencies. Items on the TEMA-3 included producing finger displays to represent different quantities, counting, and making numerical comparisons. All children started on the first item of the test and continued until they failed five consecutive items.

There was a break in the distribution of TEMA scores between the 21st and 27th national percentile ranks; thus, children with scores less than the 22nd percentile were categorized as at risk for MLD ($n = 34$) and the remaining children categorized as TA ($n = 34$). The respective mean percentile ranks on the TEMA were 9th and 56th for the MLD and TA groups [$F(1, 66) = 146.09$, $p < 0.0001$], consistent with ranks found for older children with MLD (Geary et al., 2012b). The groups differed in intelligence ($p < 0.0001$), executive functions ($p < 0.05$), and letter identification scores ($p < 0.0001$), as shown in **Table 1**; however, the difference on the TEMA remained significant when these scores were covaried (lsmeans = 13th and 51st respective percentile rank, $p < 0.0001$). In contrast, there were no group differences in level of parental education, $\chi^2(6) = 8.25$, $p = 0.2207$.

Because the TEMA-3 and our quantitative tasks are based on the same research literature, there is some overlap in the assessed competencies. For the ages assessed here, our tasks cover a broader range of competencies and include more difficult items for overlapping ones. The mean and standard deviation of the raw score of the MLD group indicated that they were, on average, successful on TEMA-3 items that involved identifying a set of up to three items, showing the examiner up to five fingers, counting to five, and identifying *more* when comparing simultaneously presented sets less than 11. The primary overlap is for our enumeration and verbal counting tasks, and in both cases our range of potential counts is higher than the items on the TEMA-3. Conceptually the *more* task overlaps with our ordinality tasks. However, there are no explicit numeral comparison items at this point in the TEMA-3,

Table 1 | Academic and ability scores across MLD status.

Test	Group	
	MLD	**TA**
Intelligence	92 (13)	104 (17)
TEMA	79 (6)	102 (9)
Executive functions	28 (12)	34 (14)
Letter identification	6 (7)	17 (9)
Parental education	3.5 (1.0)	3.6 (0.8)

Parenthetical values are standard deviations. MLD, mathematical learning disability; TA, typically achieving; TEMA, standard scores (M = 100, SD = 15) from the Test of Early Mathematical Abilities-3 (Ginsburg and Baroody, 2003); Executive functions scores are the mean number of correct conflict items with a minimum possible score of 15 and a maximum of 60.

Table 2 | Sequence of tasks and ages.

Sequence of tasks	Age of children
Quant 1 (fall)	
Enumeration	Mean: 3y9m
Give-a-number	Range: 3y2m–4y2m
Point-to-x	
Magic box	
Discrete quantity discrimination (ANS)	
Ordinal choice	
Quant 2 (fall)	
Verbal counting	Mean: 3y11m
Non-verbal calculation	Range: 3y4m–4y3m
Numeral recognition	
Numeral comparison	
Counting knowledge	
Continuous quantity discrimination	
Cognitive battery	
Executive functions (card-sorting)	Mean: 4y0m
WPPSI-III (receptive vocabulary, block design, information)	Range: 3y6m–4y5m
PALS (upper-case alphabet recognition)	
Quant 1 (spring)	Mean: 4y2m
	Range: 3y6m–4y7m
Quant 2 (spring)	Mean: 4y2m
	Range: 3y7m–4y8m
Test of early mathematics ability-3	Mean: 4y3m
	Range: 3y8m–4y8m

and our ordinal choice task involves comparisons of sets of items with sequential presentation and smaller (i.e., more difficult) ratios between sets than the TEMA-3 *more* item.

The mean and standard deviation of the TEMA-3 score of the TA group indicated that they were, on average, able to answer additional items that assessed counting up to 10, cardinal knowledge using counting and give-a-number, numeral identification, and non-verbal calculation. The three latter items are similar to our tasks, but our tasks included more items and somewhat more difficult items. For the ages assessed here, most of the children would not have been administered TEMA-3 items that overlapped with the majority of our tasks, including counting knowledge, numeral comparison, point-to-x, discrete (ANS task) and continuous magnitude, or magic box tasks.

PROCEDURE

Children were tested individually in six testing sessions lasting approximately 35 min each. All sessions were completed in their preschool facility. *Quant 1* [enumeration, give-a-number, point-to-x, magic box, discrete quantity discrimination (ANS), and ordinal choice, in that order] and *Quant 2* (verbal counting, non-verbal calculation, numeral recognition, numeral comparison, counting knowledge, and continuous quantity discrimination, in that order) were each administered, in separate sessions, once at the beginning of the fall semester and once in the middle of the spring semester. At the beginning of the spring semester, children were tested in a single session that included the EF scale (Beck et al., 2011), the WPPSI-III (Wechsler, 2002), and letter identification (Invernizzi et al., 2004). The final testing session consisted of the TEMA-3 (Ginsburg and Baroody, 2003), and was administered at the end of the spring semester. The sequence of testing and mean ages at each assessment are provided in **Table 2**. The experimental procedure was reviewed and approved by the Institutional Review Board of the University of Missouri. Written consent was obtained from all parents, and all participants provided verbal assent for all assessments.

In order to encourage children and keep them motivated during testing sessions, they received stickers after completion of tasks or blocks of trials; at the end of testing sessions, children also received educational prizes (e.g., age-appropriate books). All sessions were videotaped and the video records were used for coding and to determine reliability. Trained observers naïve to the purpose of the study reviewed testing sessions and recorded data for 15 randomly selected participants. Reliability was calculated separately for each of the 12 quantitative tasks by correlating the data collected during the test session with that recoded from the videotapes. Reliability was >0.92 for all the tasks, with one exception (time 2 Counting Knowledge = 0.86). Given the very high reliabilities, all analyses were conducted with the data collected during the testing sessions.

ANALYSES

Missing observations (6%) were estimated (maximum likelihood estimates with five imputations) using the multiple imputations program of SAS Institute (2004). To reduce the number of quantitative variables and the risk of false positives, the 12 time 1 and time 2 quantitative tasks were submitted to a principal components factor analysis (promax rotation) and factors with Eigenvalues >1 were retained (Gorsuch, 1983). The enumeration, give-a-number, numeral recognition, and verbal counting tasks loaded on the same factor at both times of measurement, and thus composite number knowledge factor scores were created using the mean of these four tasks for time 1 (α = 0.80) and time 2 (α = 0.85). In

addition, overall scores for the TEMA were regressed on the time 1 number knowledge composite along with the eight remaining individual quantitative task variables using a stepwise procedure (forward selection with $p < 0.05$ to enter and stay). We focused on time 1 because competence at the beginning of preschool is of greater practical and theoretical importance than that assessed toward the end of a year of preschool (Libertus et al., 2011). The number knowledge composite ($r^2 = 0.57$, $p < 0.0001$) was selected first, followed by ordinal choice ($pr^2 = 0.06$, $p < 0.002$); no other quantitative variables were selected. Thus, all subsequent analyses included the number knowledge composite and ordinal choice variables, along with the Weber fraction and percent correct variables from the ANS task. The latter were included based on our *a priori* prediction of poor ANS acuity for children with MLD.

Following analyses of group differences (MLD vs. TA) on the ANS task Weber fraction and percent correct variables, random intercept mixed models were run to estimate group differences on the number knowledge composite and ordinal choice variables across time 1 and time 2, using age, sex, parental education, intelligence, executive control, and letter identification scores as covariates. Logistic regressions were then used to determine the odds of MLD status using the time 1 number knowledge composite and ordinal choice variables along with ANS percent correct as predictors, using the same covariates as in the mixed models. The intelligence, executive control, and letter identification variables were standardized ($M = 0$, $SD = 1$) for all of the analyses. Regressing mathematics achievement on the Weber fraction and percent correct variables resulted in a significant relation for percent correct ($p = 0.0091$) but not the Weber variable ($p = 0.8584$). Thus, only the percent correct variable was used in the logistic regressions.

RESULTS

The first set of analyses addresses the hypothesis that children at risk for MLD have deficits in the ANS system, whereas the second set addresses group differences in rate of development of the quantitative competencies assessed by the number knowledge composite and the ordinal choice variables. In the third set, we focus on the predictive utility of the ANS percent correct, number knowledge composite, and ordinal choice variables for predicting the odds of MLD status at the end of 1 year of preschool. In the final section, we present an assessment of variability in task performance for children in the MLD group to determine if quantitative deficits are uniform or variable for these children.

MATHEMATICAL LEARNING DISABILITY AND ANS ACUITY

Median Weber fractions ($w = 0.71$, 0.59, for time 1 and time 2, respectively) and percent correct (63, 66% correct for time 1 and time 2, respectively) were consistent with previous studies of children of the same age (Libertus et al., 2011; Mazzocco et al., 2011b). However, many children had difficulty with the task and thus the Weber fractions are based on only 48 and 45 of the 68 children for time 1 and time 2, respectively. To increase the amount of useable data, we took each child's best (smallest) Weber fraction and highest percent correct across the two times of measurement,

resulting in a median Weber fraction of 0.56 ($n = 61$) with 75% correct ($n = 68$).

Children at risk for MLD had higher mean Weber fractions ($M = 1.65$, $SD = 1.61$) than their TA peers ($M = 0.84$, $SD = 1.06$), $F(1, 66) = 6.05$, $p < 0.02$, $d = 0.61$, and they were less accurate on the ANS task ($M = 69\%$, $SD = 14$) than the TA children ($M = 81\%$, $SD = 16$), $F(1, 66) = 10.67$, $p < 0.002$, $d = 0.79$. Control of group differences in intelligence, executive functions, and letter identification scores resulted in a non-significant group difference for the Weber fraction ($p = 0.2019$), but the difference in percent correct remained significant ($p = 0.029$, lsmeans = 71 and 80% for the MLD and TA groups, respectively).

As a follow-up, we created subgroups of MLD ($n = 16$) and TA ($n = 12$) children with intelligence scores between 90 and 110. These subgroups did not differ on intelligence ($M = 98$, $SD = 5$; $M = 100$, $SD = 6$ for the MLD and TA groups respectively) or executive functions ($M = 27$, $SD = 12$; $M = 33$, $SD = 13$) ($ps < 0.2303$), but they did for letter identification scores ($M = 9$, $SD = 8$; $M = 17$, $SD = 7$, $p < 0.0285$). The mean standard mathematics achievement score was 81 ($SD = 5.2$) for the MLD subgroup and 103 ($SD = 11$) for the TA subgroup ($p < 0.0001$). The MLD subgroup had a higher mean Weber fraction ($M = 1.35$, $SD = 1.57$) than the TA subgroup ($M = 0.58$, $SD = 0.75$) but the difference was not significant ($p = 0.129$). Mean percent correct was 74% ($SD = 15$) for the MLD subgroup and 87% ($SD = 13$) for the TA subgroup, which was significant ($p = 0.0227$, $d = 0.93$).

MATHEMATICAL LEARNING DISABILITY AND QUANTITATIVE DEVELOPMENT

Mean scores (percent of maximum scores) across times of measurement for the MLD and TA groups for the ordinal choice and four quantitative tasks that compose the number knowledge composite are shown in **Figure 1**. The second and third columns of **Table 3** show the summary results for group differences on the number knowledge composite. Controlling for other variables in the model, number knowledge composite accuracy was 13.8% lower at time 1 than time 2 ($p < 0.0001$) and the children at risk for MLD scored 14.1% lower than their TA peers at time 1 ($p < 0.0001$). The gap between the MLD and TA groups widened to 20% ($14.1 + 5.9$) by time 2 ($p < 0.0050$). Follow-up analyses of the four individual tasks revealed the MLD/TA contrast was significant for enumeration ($p < 0.0129$), give-a-number ($p < 0.0001$), and verbal counting ($p < 0.0001$), as was the interaction between time and MLD contrasts for enumeration ($p < 0.0155$), give-a-number ($p < 0.0296$), and verbal counting ($p < 0.0312$). For all of the latter tasks, the gap between the MLD and TA groups widened from time 1 to time 2.

The two rightmost columns of **Table 3** show that the children at risk for MLD scored 9.6% lower than their TA peers on the ordinal choice task at time 1 ($p = 0.0478$), with no significant change in this deficit across time 1 and time 2. Follow-up analyses indicated the children at risk for MLD did not score above chance (50%) for either time 1 ($t_{33} = 0.57$, $p = 0.5744$) or time 2 ($t_{33} = 0.64$, $p = 0.5241$) on the ordinal choice task, but the TA children did; time 1 ($t_{33} = 3.75$, $p = 0.0007$), time 2 ($t_{33} = 2.22$, $p = 0.0331$).

FIGURE 1 | Boxplots for the scores of the MLD (0) and TA (1) groups for the ordinal choice (the dashed line indicates 50% chance performance) and four quantitative tasks (Enumeration, **Give a number, Numeral Recognition, and Verbal Counting) that compose the number knowledge composite.** Time 1 is on top and time 2 on the bottom.
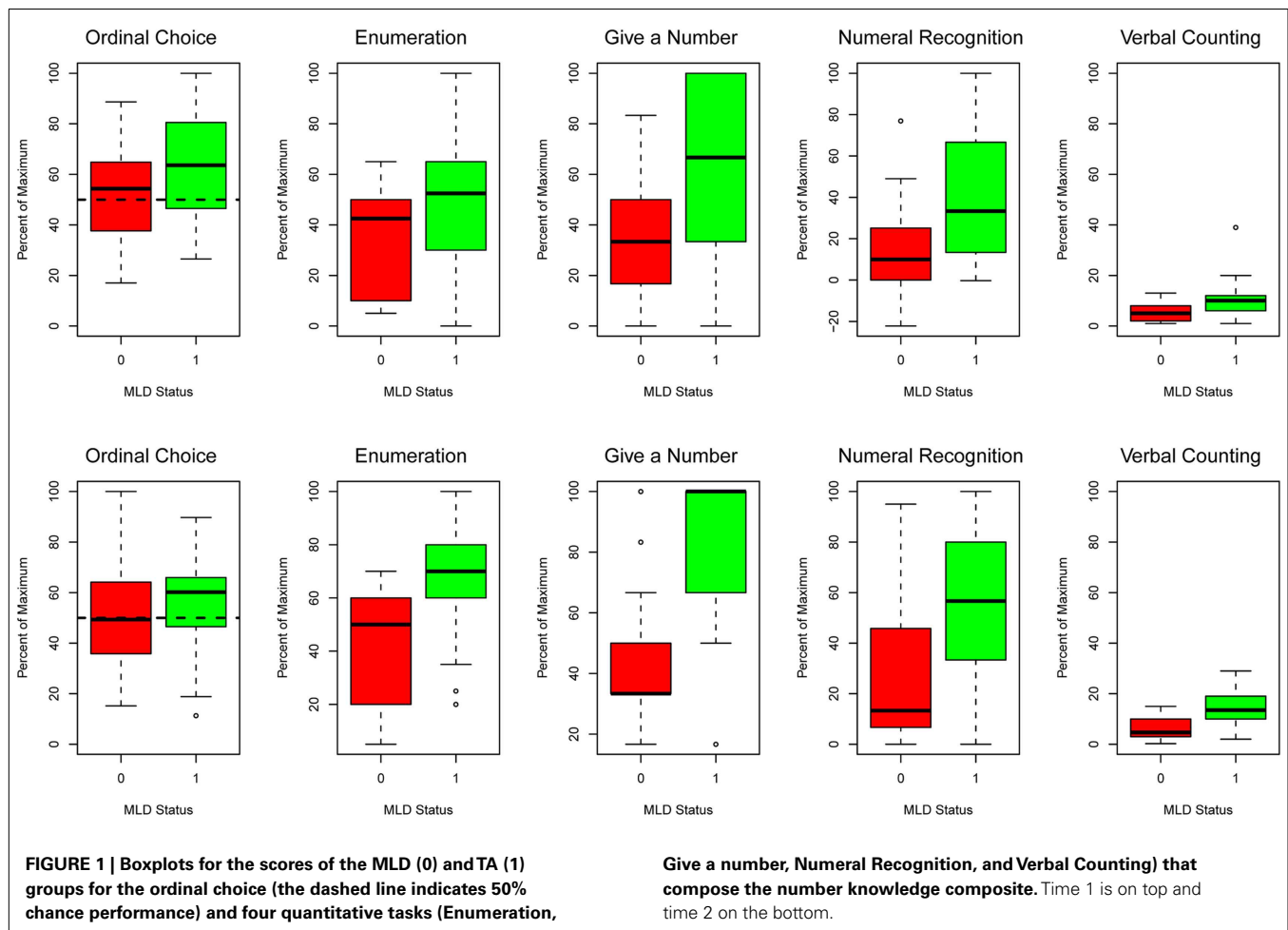
Table 3 | Mixed model results for the number knowledge composite and ordinal choice tasks.

Variable	Number knowledge composite		Ordinal choice	
	Estimate	***p***	**Estimate**	***p***
Intercept	17.0 (20.2)	0.4036	46.5 (32.3)	0.1549
Time 1 vs. time 2	−13.8 (1.47)	0.0001	5.99 (3.37)	0.0782
Age	0.71 (0.44)	0.1124	0.34 (0.71)	0.6318
Girl vs. boy	−1.35 (2.46)	0.5844	1.62 (3.93)	0.6804
NI vs. college	0.34 (3.29)	0.9180	−3.01 (5.25)	0.5681
HS vs. college	1.72 (3.06)	0.5751	−5.74 (4.89)	0.2424
Intelligence	5.10 (1.67)	0.0028	−0.38 (2.67)	0.8869
Executive functions	0.56 (1.49)	0.7067	2.22 (2.39)	0.3534
Letter identification	7.01 (1.56)	0.0001	−5.72 (2.50)	0.0235
MLD vs. TA	−14.1 (2.93)	0.0001	−9.64 (4.83)	0.0478
MLD by time	5.94 (2.08)	0.0050	−6.43 (4.77)	0.1797

Parenthetical values are standard errors. NI, no information on parental education; HS, high school; MLD, mathematical learning disability; TA, typically achieving. Negative values indicate the contrasted group (e.g., girls) had lower scores than the contrast group (e.g., boys), and positive values indicate the opposite.

QUANTITATIVE PERFORMANCE AND ODDS OF MLD STATUS

Approximate number system task percent correct, and time 1 ordinal choice and number knowledge composite scores were used in separate analyses to predict the odds of MLD status, controlling for sex, parental education, intelligence, executive functions, and letter identification scores. A 1 SD decrease in ANS task percent correct was associated with a 2.4-fold increase in the odds of being classified as at risk for MLD [$\chi^2(1) = 5.42$, $p = 0.0199$, 95% CI, 1.15–5.12]. The corresponding estimates for the ordinal choice and number knowledge composite variables were 3.6 [$\chi^2(1) = 6.77$, $p = 0.0093$, 95% CI, 1.37–9.55] and 4.5 [$\chi^2(1) = 6.04$, $p = 0.014$, 95% CI, 1.36–15.11], respectively. Simultaneously estimating all three quantitative variables produced a significant likelihood ratio for the overall model, $\chi^2(9) = 45.2$, $p < 0.0001$, and the estimates for the number knowledge composite and ordinal choice variables were significant, as shown in **Table 4**.

VARIATION WITHIN THE MLD GROUP

Group differences in mean levels of performance (**Figure 1**) suggest uniform quantitative deficits for children composing the MLD group. Previous studies, however, indicate that even within such groups there are often children who perform relatively well on some quantitative tasks (Denvir and Brown, 1986; Dowker, 2005b;

Table 4 | Estimates from logistic regression.

Variable	Estimate	p	Odds ratio	95% CI
Intercept	−3.16 (1.38)	0.0219	–	–
Girl vs. boy	0.49 (0.41)	0.2292	2.7	0.53–13.29
NI vs. college	−0.19 (0.57)	0.7376	2.4	0.32–17.03
HS vs. college	1.24 (0.67)	0.0637	9.8	0.99–97.89
Intelligence	−0.16 (0.50)	0.7539	0.9	0.32–2.29
Executive functions	−0.71 (0.48)	0.1386	0.5	0.19–1.26
Letter identification	1.54 (0.56)	0.0056	4.7	1.57–13.96
ANS percent correct	0.28 (0.49)	0.5681	1.3	0.51–3.43
Ordinal choice	1.40 (0.55)	0.0111	4.1	1.38–12.00
Number composite	1.96 (0.83)	0.0189	7.1	1.38–36.46

Parenthetical values are standard errors. NI, no information on parental education; HS, high school; MLD, mathematical learning disability; TA, typically achieving. Negative values indicate the contrasted group (e.g., girls) had lower scores than the contrast group (e.g., boys), and positive values indicate the opposite. The odds ratio = $e^{estimate}$.

Jordan et al., 2009a). To assess this possibility, we first categorized performance on the ANS task percent correct and on the five tasks (across both times of measurement) shown in **Figure 1** as above or below the overall mean (across both groups). Children in the MLD group scored above the mean on an average of 3.5 (range 0–8) of the 11 tasks, as compared to 7.5 (range 2–11) tasks for the TA group [$\chi^2 (11) = 31.8, p = 0.0008$]. Only four of the 34 children at risk for MLD scored below average on all 11 tasks. Fifteen of these 34 children had an above average percent correct on the ANS task, and 12 of them scored above average across both times of measurement for one of the five tasks shown in **Figure 1**; five children scored above average on two tasks and one child on three tasks.

These did not substantially influence group-level mean scores, because different children within the MLD group tended to perform well on different tasks. Three scored above average for both times of measurement for the give-a-number task, four for the numeral recognition and verbal counting tasks, six for the ordinal choice task, and eight for the enumeration task.

DISCUSSION
The study of children enrolled in Title I preschool is well suited to our goal of identifying early risks of later MLD and LA. Indeed, after a year of preschool, half of the children in our sample had mathematics achievement scores in the same range as those found in school-age children with MLD or LA (Geary et al., 2007; Murphy et al., 2007). It is premature to consider these preschoolers as MLD or LA, but they appear to be at high risk of becoming so, independent of the effects of intelligence, executive functions, and parental education on mathematics achievement. The high proportion of children at risk for MLD or LA allowed us to test the hypothesis that an impaired ANS is the core deficit underlying their poor mathematics achievement (Piazza et al., 2010), and the broader assessment of quantitative competencies allowed us to gauge the importance of the ANS relative to other early competencies that are predictive of later mathematics achievement (Locuniak and Jordan, 2008; Jordan et al., 2009b).

The results provide some support for the hypothesis that an impaired ANS contributes to the mathematics achievement deficits of children at risk for MLD (Piazza et al., 2010; Mazzocco et al., 2011b), but the results are not definitive. The at risk group had, as predicted, higher Weber fractions (i.e., less fidelity) and lower accuracy on the ANS task. Moreover, a 1 SD decrease in ANS task percent correct resulted in a substantial increase in the odds (2.4) of being classified as at risk for MLD after a year of preschool. However, controlling for intelligence, executive control, and preliteracy (letter identification) scores eliminated the group difference for the Weber fraction and attenuated it for percent correct.

On the one hand, these analyses suggest that control of other factors that might affect mathematics learning (intelligence, executive control) is important for testing the ANS hypothesis. On the other hand, Piazza et al. (2010) controlled for intelligence and Mazzocco et al. (2011a) controlled for working memory and still found ANS deficits in school-age children with MLD. One source of the across-study discrepancies is in the characteristics of the samples defined as MLD, or dyscalculic (Piazza et al., 2010). For instance, our sample of at risk children had low average intelligence scores, as is commonly found (Geary et al., 2007), but the children identified as dyscalculic in the Piazza et al. study were of average intelligence. Indeed, our analyses of MLD and TA subgroups matched on intelligence confirmed Piazza et al.'s findings but the results were only significant for ANS task percent correct and not the Weber fraction. No doubt our small sample sizes for these subgroups were a contributing factor. As we recruit additional children into the study, we will be able to obtain a larger sample of preschool children of average intelligence and effortful control and very low mathematics achievement scores and will then be able to provide a more sensitive replication attempt of the Piazza et al. (2010) and Mazzocco et al. (2011a) findings.

Regardless, simultaneous estimation of group differences on the number knowledge composite and ordinal choice tasks eliminated the significance of ANS task percent correct in predicting the odds of MLD status. Thus, for children who are entering preschool the best predictors of risk for MLD, independent of intelligence, executive control, preliteracy scores, and parental education, is poor knowledge of Arabic numerals, number words, and their cardinal values. These children not only began preschool behind their TA peers on these tasks, they fell further behind as the year progressed. These findings do not mean poor ANS fidelity did not contribute to these children's low mathematics achievement. We suggest that any such deficit may largely operate through ease of learning the relation between Arabic numerals, number words and the magnitudes they represent. This is consistent with Rousselle and Noël's (2007) hypothesis that mapping symbols onto magnitude representations contributes to the deficits of children with MLD, but further suggests that the fidelity of ANS representations themselves may influence the mapping process.

An unexpected finding was that the children at risk for MLD were unable, even at the end of a year of preschool, to discriminate more and less on the ordinal choice task. Performance on this task is not dependent on an understanding of numerals or number words and may be dependent on properties of the ANS that are not captured by the discrete discrimination task (Gallistel and Gelman, 1992; Gallistel, 2007). Gallistel and Gelman (2005)

argued the most important aspect of the ANS is that the generated magnitude representations embody cardinality and ordinality information. However, it will require a larger sample of preschoolers to fully explore whether or not these at risk children's deficit on this task is related to poor ANS acuity or other properties of this system.

Finally, our results confirm previous studies that have shown that even within MLD/LA groups many children will show normal or better performance in some quantitative areas (Denvir and Brown, 1986; Geary, 1990; Geary et al., 1991) and more generally that the development of quantitative competencies is uneven (Dowker, 2005b; Jordan et al., 2009a). Previous studies of MLD/LA school age children suggest that those with at least some intact quantitative competencies show larger across-grade achievement gains than their peers with deficits in multiple areas (Geary et al., 1991). These results suggest that the children in our at risk group who showed multiple islets of normal performance on the quantitative tasks may not in fact be MLD in the long-term. Follow-up of these children will determine if this is in fact that case. Either way, our results and related ones indicate that many MLD/LA children will have some quantitative strengths that may be potential building blocks for remedial interventions.

CONCLUSION

Preschoolers with a strong intuitive sense of quantity, as measured by their ability to quickly determine which of two collections of objects has more (ANS task), score higher on mathematics achievement tests than other children, controlling for intelligence, effortful control, and preliteracy knowledge. Preschoolers at high risk for a learning disability in mathematics have a poor intuitive sense of quantity, but their poor understanding of more and less, and slow learning of Arabic numerals, number words, and their meanings may constitute a stronger long-term risk.

ACKNOWLEDGMENTS

REFERENCES

Barbaresi, W. J., Katusic, S. K., Colligan, R. C., Weaver, A. L., and Jacobsen, S. J. (2005). Math learning disorder: incidence in a population-based birth cohort, 1976–1982, Rochester, Minn. *Ambul. Pediatr.* 5, 281–289.

Beck, D. M., Schaefer, C., Pang, K., and Carlson, S. M. (2011). Executive function in preschool children: test-retest reliability. *J. Cogn. Dev.* 12, 169–193.

Berch, D. B., and Mazzocco, M. M. M. (eds). (2007). *Why is Math So Hard for Some Children? The Nature and Origins of Mathematical Learning Difficulties and Disabilities.* Baltimore, MD: Paul H. Brookes Publishing Co.

Blair, C., and Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Dev.* 78, 647–663.

Blatchford, R., Burke, J., Farquhar, C., Plewis, I., and Tizard, B. (1987). Associations between pre-school reading related skills and later reading achievement. *Br. Educ. Res. J.* 13, 15–23.

Briars, D., and Siegler, R. S. (1984). A featural analysis of preschoolers' counting knowledge. *Dev. Psychol.* 20, 607–618.

Bugden, S., and Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition* 118, 32–44.

Bull, R., Espy, K. A., and Wiebe, S. A. (2008). Short-term memory, working memory, and executive functions in preschoolers: longitudinal predictors of mathematical achievement at age 7 years. *Dev. Neuropsychol.* 33, 205–228.

Butterworth, B. (2005). "Developmental dyscalculia," in *Handbook of Mathematical Cognition*, ed. J. I. D. Campbell (Hove: Psychology Press), 455–467.

Bynner, J. (1997). Basic skills in adolescents' occupational preparation. *Career Dev. Q.* 45, 305–321.

Clark, C. A. C., Pritchard, V. E., and Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Dev. Psychol.* 46, 1176–1191.

De Smedt, B., and Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *J. Exp. Child Psychol.* 108, 278–292.

Deary, I. J., Strand, S., Smith, P., and Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence* 35, 13–21.

Denvir, B., and Brown, M. (1986). Understanding of number concepts in low attaining 7–9 year olds: part I. Development of descriptive framework and diagnostic instrument. *Educ. Stud. Math.* 17, 15–36.

Dowker, A. (2005a). *Individual Differences in Arithmetic: Implications for Psychology, Neuroscience and Education.* Hove: Psychology Press.

Dowker, A. (2005b). Individual differences in numerical abilities in preschoolers. *Dev. Sci.* 11, 650–654.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Dev. Psychol.* 43, 1428–1446.

Feigenson, L., Carey, S., and Hauser, M. (2002). The representations underlying infants' choice of more: object-files versus analog magnitudes. *Psychol. Sci.* 13, 150–156.

Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci. (Regul. Ed.)* 8, 307–314.

Gallistel, C. R. (2007). Commentary on Le Corre & Carey. *Cognition* 105, 439–445.

Gallistel, C. R., and Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition* 44, 43–74.

Gallistel, C. R., and Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends Cogn. Sci. (Regul. Ed.)* 4, 59–65.

Gallistel, C. R., and Gelman, R. (2005). "Mathematical cognition," in *The Cambridge Handbook of Thinking and Reasoning*, Vol. 23, eds K. Holyoak and R. Morrison (Cambridge: Cambridge University Press), 559–588.

Geary, D. C. (1990). A componential analysis of an early learning deficit in mathematics. *J. Exp. Child Psychol.* 49, 363–383.

Geary, D. C. (2011). Cognitive predictors of individual differences in achievement growth in mathematics: a five year longitudinal study. *Dev. Psychol.* 47, 1539–1552.

Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Curr. Dir. Psychol. Sci.* 22, 23–27.

Geary, D. C., Bow-Thomas, C. C., and Yao, Y. (1992). Counting knowledge and skill in cognitive addition: a comparison of normal and mathematically disabled children. *J. Exp. Child Psychol.* 54, 372–391.

Geary, D. C., Brown, S. C., and Samaranayake, V. A. (1991). Cognitive addition: a short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Dev. Psychol.* 27, 787–797.

Geary, D. C., Hoard, M. K., and Bailey, D. H. (2012a). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *J. Learn. Disabil.* 45, 291–307.

Geary, D. C., Hoard, M. K., Nugent, L., and Bailey, D. H. (2012b). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: a five year prospective study. *J. Educ. Psychol.* 104, 206–223.

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., and Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Dev.* 78, 1343–1359.

Geary, D. C., Hoard, M. K., Nugent, L., and Bailey, H. D. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS ONE* 8:e54651. doi:10.1371/journal.pone.0054651

Gelman, R., and Gallistel, C. R. (1978). *The Child's Understanding of Number.* Cambridge, MA: Harvard University Press.

Gilmore, C. K., McCarthy, S. E., and Spelke, E. S. (2010). Non-symbolic arithmetic skills and mathematics achievement in the first year of formal schooling. *Cognition* 115, 394–406.

Ginsburg, H. P., and Baroody, A. J. (2003). *Test of Early Mathematical Ability,* 3rd Edn. Austin, TX: Pro-Ed.

Gorsuch, R. L. (1983). *Factor Analysis,* 2nd Edn. Hillsdale: Lawrence Erlbaum Associates.

Halberda, J., and Feigenson, L. (2008). Developmental change in the acuity of the "number sense": the approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Dev. Psychol.* 44, 1457–1465.

Halberda, J., Mazzocco, M., and Feigenson, L. (2008). Individual differences in nonverbal number acuity correlate with maths achievement. *Nature* 455, 665–669.

Hannula, M. M., Lepola, J., and Lehtinen, E. (2010). Spontaneous focusing on numerosity as a domain-specific predictor of arithmetical skills. *J. Exp. Child Psychol.* 107, 394–406.

Invernizzi, M., Sullivan, A., Meier, J., and Swank, L. (2004). *PreK Teacher's Manual: Phonological Awareness and Literacy Screening.* Charlottesville, VA: University of Virginia.

Iuculano, T., Tang, J., Hall, C. W. B., and Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Dev. Sci.* 11, 669–680.

Jordan, J.-A., Mulhern, G., and Wylie, J. (2009a). Individual differences in trajectories of arithmetical development in typically achieving 5- to 7-year olds. *J. Exp. Child Psychol.* 103, 455–468.

Jordan, N. C., Kaplan, D., Ramineni, C., and Locuniak, M. N. (2009b). Early math matters: kindergarten number competence and later mathematics outcomes. *Dev. Psychol.* 45, 850–867.

Koontz, K. L., and Berch, D. B. (1996). Identifying simple numerical stimuli: processing inefficiencies exhibited by arithmetic learning disabled children. *Math. Cogn.* 2, 1–23.

Krajewski, K., and Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: findings from a 3-year longitudinal study. *J. Exp. Child Psychol.* 103, 516–531.

Landerl, K., Bevan, A., and Butterworth, B. (2003). Developmental dyscalculia and basic numerical capacities: a study of 8-9-year-old students. *Cognition* 93, 99–125.

Le Corre, M., and Carey, S. (2007). One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition* 105, 395–438.

LeFevre, J.-A., Fast, L., Skwarchuk, S.-L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: longitudinal predictors of performance. *Child Dev.* 81, 1753–1767.

Levine, S. C., Huttenlocher, J., and Jordan, N. (1992). Development of calculation abilities in young children. *J. Exp. Child Psychol.* 53, 72–93.

Lewis, C., Hitch, G. J., and Walker, P. (1994). The prevalence of specific arithmetic difficulties and specific reading difficulties in 9-year-old to 10-year-old boys and girls. *J. Child Psychol. Psychiatry* 35, 283–292.

Libertus, M. E., Halberda, J., and Feigenson, L. (2011). Preschool acuity of the approximate number system correlates with math abilities. *Dev. Sci.* 14, 1292–1300.

Locuniak, M. N., and Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *J. Learn. Disabil.* 41, 451–459.

Lyons, I. M., and Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition* 121, 256–261.

Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011a). Impaired acuity of the approximate number system underlies mathematical learning disability (Dyscalculia). *Child Dev.* 82, 1224–1237.

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

Morgan, P. L., Farkas, G., and Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *J. Learn. Disabil.* 42, 306–321.

Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., and Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *J. Learn. Disabil.* 40, 458–478.

Piazza, M., Facoetti, A., Trussardi, A. M., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41.

Rivera-Batiz, F. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *J. Hum. Resour.* 27, 313–328.

Rousselle, L., and Noël, M.-P. (2007). Basic numerical skills in children with mathematical learning disabilities: a comparison of symbolic vs. non-symbolic number magnitude processing. *Cognition* 102, 361–395.

Sarnecka, B. W., and Carey, S. (2008). How counting represents number: what children must learn and when they learn it. *Cognition* 108, 662–674.

SAS Institute. (2004). *Statistical Analysis System 7.5.* Cary, NC: SAS Institute.

Shalev, R. S., Manor, O., and Gross-Tsur, V. (2005). Developmental dyscalculia: a prospective six-year follow-up. *Dev. Med. Child Neurol.* 47, 121–125.

Starkey, P. (1992). The early development of numerical reasoning. *Cognition* 43, 93–126.

Strauss, M. S., and Curtis, L. E. (1984). "Development of numerical concepts in infancy," in *Origins of Cognitive Skills: The Eighteenth Annual Carnegie Symposium on Cognition,* ed. C. Sophian (Hillsdale, NJ: Erlbaum), 131–155.

vanMarle, K. (2013). Infants use different mechanisms to make small and large number ordinal judgments. *J. Exp. Child Psychol.* 114, 102–110.

vanMarle, K., Aw, J., McCrink, K., and Santos, L. R. (2006). How capuchin monkeys (*Cebus apella*) quantify objects and substances. *J. Comp. Psychol.* 120, 416–426.

vanMarle, K., and Wynn, K. (2001). "Preschoolers' detection of magical events," in *Poster Presented at the 2001 Meeting of the Society for Research in Child Development,* Minneapolis, MN.

vanMarle, K., and Wynn, K. (2011). Tracking and quantifying objects and non-cohesive substances. *Dev. Sci.* 14, 302–316.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence-III.* San Antonio, TX: The Psychological Corporation, Harcourt Brace & Co.

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

Development of numerical processing in children with typical and dyscalculic arithmetic skills—a longitudinal study

Karin Landerl *

Department of Psychology, University of Graz, Graz, Austria

Numerical processing has been demonstrated to be closely associated with arithmetic skills, however, our knowledge on the development of the relevant cognitive mechanisms is limited. The present longitudinal study investigated the developmental trajectories of numerical processing in 42 children with age-adequate arithmetic development and 41 children with dyscalculia over a 2-year period from beginning of Grade 2, when children were 7; 6 years old, to beginning of Grade 4. A battery of numerical processing tasks (dot enumeration, non-symbolic and symbolic comparison of one- and two-digit numbers, physical comparison, number line estimation) was given five times during the study (beginning and middle of each school year). Efficiency of numerical processing was a very good indicator of development in numerical processing while within-task effects remained largely constant and showed low long-term stability before middle of Grade 3. Children with dyscalculia showed less efficient numerical processing reflected in specifically prolonged response times. Importantly, they showed consistently larger slopes for dot enumeration in the subitizing range, an untypically large compatibility effect when processing two-digit numbers, and they were consistently less accurate in placing numbers on a number line. Thus, we were able to identify parameters that can be used in future research to characterize numerical processing in typical and dyscalculic development. These parameters can also be helpful for identification of children who struggle in their numerical development.

Keywords: dyscalculia, numerical processing development, number comparison, dot counting, number line

Efficient processing of numbers and numerical sets in young children has been found to predict later arithmetic skills (Mazzocco and Thompson, 2005; Halberda and Feigenson, 2008; de Smedt et al., 2009; Jordan et al., 2009, 2010; Geary, 2011). There is also converging evidence that numerical processing is deficient in individuals with dyscalculia, a severe and persistent disability in learning arithmetic which can be highly selective, affecting learners with normal intelligence (Butterworth et al., 2011). Basic numerical processing has been proposed to constitute an innate core mechanism which is evident in infants (Xu and Spelke, 2000; Xu and Arriaga, 2007) and underlies all further developments in number processing (Butterworth, 1999; Wilson and Dehaene, 2007; Dehaene, 2011).

Although an association between numerical processing and arithmetic is clearly established, the construct of numerical processing itself is still underspecified: First, various different tasks (shortly described in the next section) have been used to investigate how humans represent and process numbers and numerical sets in their cognitive system. It is as yet unclear which tasks and parameters are best suited to measure typical and atypical developmental trajectories within the domain of numerical processing. Second, up to date, empirical evidence on the development of basic numerical skills comes mostly from cross-sectional studies (Girelli et al., 2000; Holloway and Ansari, 2008; Landerl and

Kölle, 2009; Schleifer and Landerl, 2011). In the current study, we repeatedly presented a battery of numerical processing tasks to children with good and poor arithmetic skills during their elementary school years (Grades 2–4), allowing a detailed view on developmental processes. Before explicating the outline of the current study in detail, we will give an overview of the tasks and effects used by previous studies to assess numerical development.

In the dot enumeration paradigm, participants have to count a limited number of dots (usually no more than 10) as quickly as possible. The efficiency of counting procedures increases over time (e.g., Jordan et al., 2006; Reeve et al., 2012). Enumeration tasks induce a characteristic pattern of performance, indicating two distinct enumeration systems (Vetter et al., 2011): small numerosities up to three or four are typically responded to with high accuracy and speed. This process of rapid identification of small dot numbers is termed *subitizing*. When counting higher numerosities, reaction times and error rates rise with increasing numerosity, indicating the execution of a sequential counting procedure. In a recent cross-sectional study, Schleifer and Landerl (2011) found adult-like subitizing performance in 11-year old, but not in younger children. Full competence in sequential counting of larger dot arrays was only evident in 14-year olds, while younger age groups performed at less proficient levels. The only study that assessed dot counting performance

longitudinally (7 assessments between the ages of 6 and 11 years) also reported a consistent decrease of response times with increasing age as well as a growing subitizing range (Reeve et al., 2012). While 6-year-old children typically subitized two dots, they were able to subitize three dots by the age of 9. A subitizing range of four dots was not achieved throughout the study. Interestingly, both, Reeve et al. (2012) as well as Schleifer and Landerl (2011) found specific subitizing problems (steeper response time slopes) in poor achievers, while in the counting range, responses were generally slower, but the gradients of response time slopes were similar across achievement groups. These findings suggest that problems in subitizing may be a particularly useful marker of dyscalculia (Butterworth, 1999).

Another simple experimental paradigm that is highly informative with respect to the cognitive representation of number is number comparison. Individuals are asked to select the numerically larger of two numbers or numerical sets (e.g., dot arrays). The speed with which this decision is made depends on the numerical distance between the two numerosities. The smaller this distance, the slower (and less accurately) the decision is made due to a larger internal overlap between the two internal magnitude representations. The acuity of non-symbolic quantity processing increases during development, allowing children to discriminate similar numerical sets more precisely (Halberda and Feigenson, 2008; Piazza et al., 2010). A symbolic distance effect has been demonstrated even among kindergarteners (Sekuler and Mierkiewicz, 1977). Acuity of non-symbolic quantity processing in kindergarten was found to predict arithmetic competence at age six (Mazzocco et al., 2011b) and interindividual differences in the acuity of quantity processing were found to be directly related to arithmetic competence (Libertus et al., 2011). Similarly, Holloway and Ansari (2008) reported a relatively smaller symbolic distance effect in higher grade levels and interpreted this age-related decrease as continuing specification of the cognitive representation of number. In line with this assumption, de Smedt et al. (2009) demonstrated an association between the symbolic distance effect and individual differences in math achievement one year later: Children with a relatively smaller distance effect in grade one had higher math scores in grade two. However, other studies found a rather stable influence of numerical distance on symbolic number comparison across different age or achievement groups, accompanied by a general decrease in response times (Girelli et al., 2000; Landerl and Kölle, 2009; Reeve et al., 2012).

Findings on symbolic and non-symbolic number comparison in dyscalculia are mixed. There is evidence for specific problems in non-symbolic magnitude comparison among dyscalculic individuals (Landerl et al., 2009; Piazza et al., 2010; Kucian et al., 2011; Mazzocco et al., 2011a), however, in some studies the deficits of dyscalculic individuals were limited to symbolic processing of Arabic numbers and did not extend to non-symbolic magnitudes (Rousselle and Noël, 2007; Iuculano et al., 2008; Landerl and Kölle, 2009). Based on this discrepancy, Rousselle and Noël (2007) have suggested that the innate core system of analog magnitude representations in itself may be intact in dyscalculia, but cannot be efficiently accessed from symbolic representations of numbers.

The number comparison paradigm has also been used to investigate the automaticity of numerical processing. When individuals are asked to decide which of two digits is physically larger, numerical value interferes with their physical judgments. Generally, incongruent items (e.g., 4 9) are responded to more slowly than congruent items (e.g., 4 9; Girelli et al., 2000; Landerl and Kölle, 2009; Bugden and Ansari, 2011). This *size-congruity effect* indicates automatic processing of numbers and requires a certain amount of experience. Cross-sectional studies show interference between physical and numerical size even in first grade (Rubinsten et al., 2002), while in other studies it was not even found in fourth graders (Landerl et al., 2004). Interindividual differences in the degree of automatization and differences in task format make it difficult to compare findings across age groups. A longitudinal design can control for such differences. During development, the size-congruity effect can be expected to become larger as a sign of increasing automatization of numerical processing. Children with dyscalculia are likely to show automatization of numerical processing to a lesser degree or at least at a later developmental stage. Indeed, earlier studies found no (Landerl and Kölle, 2009) or a reduced (Rubinsten and Henik, 2006) size-congruity effect in children with dyscalculia. However, Bugden and Ansari (2011) did not find a correlation of the size-congruity effect with children's arithmetic skills and concluded that automatic processing of numbers is not related to mathematical competence.

Number comparison paradigms with two-digit numbers have been shown to induce a distance effect as well as a compatibility effect (Nuerk et al., 2004). Thus, response accuracy and speed are generally lower when both tens and units are higher in one number (e.g., 83_62, 8 > 6, and 3 > 2) than when tens and units of the two numbers are incompatible (e.g., 82_63, 8 > 6, but 2 < 3). The *compatibility effect* indicates that multi-digit numbers are not processed holistically, but require adequate integration of the composite numerals and their place-value. Acquisition of the place-value system of Arabic numbers is an important step in the development of numerical competences (Mann et al., 2012). Accordingly, first evidence indicates that the compatibility effect is especially marked in young and unexperienced children (Landerl and Kölle, 2009; Pixner et al., 2009) and predicts later arithmetic skills (Moeller et al., 2011). Landerl and Kölle (2009) provided first evidence that the integration of two-digit numbers may pose a particular problem for children with dyscalculia.

A dominant view of the cognitive representation of number is the mental number line model, which postulates that internal representations of numbers and quantities are organized spatially from left to right (Dehaene, 2011). The formation of such a mental number line constitutes a vital step in the development of mathematical skills (Von Aster and Shalev, 2007). In order to investigate the format of these mental representations, children are asked to place particular numbers on lines with endpoints of 0 and 100, respectively, 0 and 1000. A standard finding is that young children overestimate the numerical size of small numbers, inducing a logarithmic number line function. With increasing experience, children's estimates become more realistic, shifting the function from a logarithmic to a linear curve (Siegler and Opfer, 2003; Siegler and Booth, 2004; Booth and Siegler, 2006; Berteletti et al., 2010; but see Ebersbach et al.,

2008; Moeller et al., 2009; Barth and Paladino, 2011 for different interpretations). Early competence in the number line task predicts later arithmetic skills (Geary, 2011), and benefits from number line trainings on children's mental representation of number and their arithmetic competence have been demonstrated (Siegler and Ramani, 2009; Kucian et al., 2011). Children with dyscalculia have been found to be less precise in their estimations and there is some indication that this is partly due to a delay in the logarithmic-linear shift (Landerl et al., 2009).

To sum up, a variation of different paradigms has been used to investigate the development of the cognitive representation of number. The general pattern is that the representational system of numbers and numerosities becomes more precise and more efficient during typical development, while this is not the case (at least not to the same extent) in dyscalculic individuals. However, integrating findings from different studies is often problematic due to variations in methodology and sample selection criteria. As most evidence on the development of numerical competencies comes from cross-sectional studies, current knowledge on the stability of task performance across time is scarce. Only Reeve et al. (2012) report reasonable stability of dot enumeration and symbolic number comparison. In latent cluster analyses, children were categorized into slow, medium, and fast subgroups based on their task-specific response times. Over seven assessments carried out between the ages of 6 and 11 years, 69% of a random sample of 159 children remained in the same cluster subgroup and no child changed from the medium or fast groups to the slow group. Still, this finding implies that almost one third of the sample did change subgroup at least once. Ordinal correlations of task-specific group membership at the different assessment points were significant (with the exception of number comparison at the age of 6, which may have overstrained some children), but mostly below 0.7 before the age of nine and a half.

The current study also aimed at investigating developmental trajectories of numerical processing and their stability over a longer time period. In order to get a broader picture of numerical development, we decided to use a range of standard numerical tasks that are assumed to tap into different aspects of numerical cognition. As we were particularly interested in differences in numerical development between children with typical arithmetic development and children with dyscalculia, we started our study at the end of grade 1 and selected participants based on their maths performance after 1 year of formal teaching and during the following 2 years. Children with typical arithmetic development and children with marked and persistent problems in this domain were followed longitudinally over a 2-year period and performed the numerical task battery five times throughout the study.

In particular, our research questions were: (1) How do standard effects of numerical processing develop in children with age-adequate arithmetic skills? (2) Are children with dyscalculia different from typically developing children in all numerical processing tasks, or is there a dyscalculia-specific profile? (3) Is the developmental trajectory of dyscalculia mostly delayed or are there characteristic deficiencies that cannot be explained by a general slowness in the acquisition of arithmetic skills? (4) How

stable is the development of numerical processing during the elementary school years?

MATERIALS AND METHODS
DESIGN
The present longitudinal study investigated the development of childrens' numerical skills from the beginning of Grade 2 to the beginning of Grade 4. Based on a screening at the end of Grade 1 and their performance during the study period, children were allocated to a group with age-adequate and a group with atypically poor arithmetic development. Two times per school year (October and March), children were individually tested with a computerized battery of numerical tasks, resulting in five assessment points altogether (t1–t5). Results from the screening period at the end of Grade 1 and the first individual assessment two months later at the beginning of Grade 2 are reported as one assessment point (t1).

PARTICIPANTS
The 83 participants of the current analysis (42 children with age-adequate arithmetic development and 41 children with dyscalculia) were selected as follows: during a screening period, a classroom test of arithmetic (Haffner et al., 2005) was given to 505 children at the end of first grade attending 19 different elementary schools in a south-western area of Germany. All children who performed more than 1 SD below age expectations on this test were invited for further assessment. For each child with poor arithmetic skills, another child was randomly selected from the same classroom who did not show any particular arithmetic problems (test performance not more than 0.5 SDs below the age norm). In order to rule out more general learning problems, the following exclusion criteria were applied:

– performance more than 1 SD below the age norm on a standardized test of nonverbal IQ (CFT 1; Cattell et al., 1997).
– performance more than 1 SD below age norm on a test of verbal short-term memory (German version of WISC IV digit span forward and backwards, Petermann and Petermann, 2008).
– clinical diagnosis of ADHD or performance more than 1.5 SDs below age norm on a standardized attention test (KITAP, Zimmermann et al., 2002) in order to avoid as far as possible confounds from comorbid attentional deficits.
– performance more than 2 SDs below the age norm on a standardized test of word reading (SLS 1–4, Mayringer and Wimmer, 2003) administered at the beginning of Grade 2 in order to avoid as far as possible confounds from comorbid dyslexia. Current evidence (e.g., Landerl et al., 2009) suggests that the cognitive deficits typically associated with dyslexia are largely independent from numerical processing, therefore, we decided to apply a rather conservative exclusion criterion for reading deficits.

Altogether, 139 children were followed over the whole study period. During the study period, the standardized test of arithmetic (HRT 1–4, see below) was given three times, i.e., end of Grade 1 and beginning of Grades 3 and 4. The two groups reported here were selected from the full longitudinal sample

based on the following criteria: Children with age-adequate arithmetic development had to show at least average performance (not more than 0.5 SD below age norms) in all three arithmetic assessments. Children with dyscalculia showed persistent problems in arithmetic during the whole study period. At t1, all children of this group performed more than 1 SD below the age norm. At the latter two assessment points, performance was never better than 0.5 SD below the age norm. **Table 1** shows that at all three assessment points, the average performance of the dyscalculia group was markedly deficient with about 1.5 SDs below age norm[1].

Table 1 | Participants' details.

	Typically developing (*n* = 42) Mean (*SD*)	Dyscalculic (*n* = 41) Mean (*SD*)	*d*
Percentage boys	57.10	39.00	
Age (months)	91.42 (3.80)	90.07 (4.75)	
Nonverbal IQ	115.57 (11.91)	101.93 (9.45)	1.08**
Verbal working memory[a]	11.02 (2.27)	9.83 (1.95)	0.54*
Attention[b]	49.24 (4.29)	45.19 (3.42)	0.93**
Reading[c]	109, 69 (14.02)	90.37 (14.89)	1.11**
ARITHMETIC[b]			
T1	55.21 (7.76)	34.27 (5.32)	3.15**
T3	57.4 (6.5)	37.32 (4.95)	3.48**
T5	56.17 (6.22)	36.9 (4.72)	3.49**
ADDITION[d]			
T1	17.6 (3.39)	8.83 (3.03)	2.73**
T3	24.74 (3.8)	16.15 (2.74)	2.60**
T5	29.12 (3.05)	20.63 (3.48)	2.60**
SUBTRACTION[d]			
T1	16.48 (4.31)	7.54 (3.58)	2.26**
T3	24.67 (4.62)	13.61 (4.32)	2.07**
T5	28.83 (3.35)	18.41 (4.59)	2.65**
MULTIPLICATION[d]			
T3	20.29 (4.5)	11.24 (4.73)	1.96**
T5	25.17 (4.56)	15.32 (3.98)	2.30**
DIVISION[d]			
T3	18.07 (5.26)	6.98 (4.84)	2.20**
T5	24.93 (4.95)	12.32 (5.68)	2.37**

[a] Standard Score (M: 10 / SD: 3).
[b] t-Score (M: 50 / SD: 10).
[c] Reading Quotient (M: 100 / SD: 15).
[d] Raw score (number of correct responses given in 2 min).
*p < 0.05; **p < 0.01.

---

[1] Ten children of the dyscalculia group performed more than 1.5 SDs below norm on all three assessment points, four children's performance was more than 1.5 SDs below norm on two assessments of arithmetic skills and 15 children performed more than 1.5 SDs below norm at least once during the study period. Only seven children of the dyscalculia group showed persistent but somewhat milder arithmetic deficits with performance lower than 1 SD below norm at the end of first grade and lower than 0.5 SDs below norm at the later assessment points.

TASKS

Standardized tests

Arithmetic. A standardized test of arithmetic skills, the subtests of the subscale "arithmetic operations" of the Heidelberger Rechentest (HRT 1–4) (Haffner et al., 2005), was given three times during the project period (t1: end of Grade 1, t3: beginning of Grade 3 and t5: beginning of Grade 4). At t1, children's competence in mental calculation (addition and subtraction) was assessed by specific subtests requiring children to write down as many correct answers as possible to a list of calculations (gradually increasing in difficulty) within a time limit of 2 min. Two further subtests had a slightly more complex format, but with the same 2 min time restriction (e.g., "__ −2 = 6" – supply: "4"; "9 + 1 __ 11" – supply "<"). At the later assessment points, multiplication and division were also assessed by 2 min subtests. The dependent measure was the number of correct responses combined for all subtests.

Nonverbal IQ. The CFT1 (Cattell et al., 1997) was given at the end of first Grade (t1). This test is based on Cattell's Culture Free Intelligence Test, Scale 1 (Cattell, 1950) and consists of five subtests (Substitutions, Mazes, Classifications, Similarities, and Matrices).

Working memory. The subtest *digit span* (forward and backwards) of the German version of the Wechsler Intelligence Scale for Children (Petermann and Petermann, 2008) was given during the screening period at the end of Grade 1.

Attention. A standardized computer-test battery assessing different aspects of attention (Zimmermann et al., 2002) was carried out in the middle of Grade 2 (t3). Five subtests measured children's alertness, attentional flexibility, distractibility, sustained attention, and divided attention.

Reading. At t1, a standardized reading test (Mayringer and Wimmer, 2003) was given in which children had to silently read simple sentences and mark whether the content of the sentence was right or wrong. The main criterion is reading speed, more specifically the number of correctly marked sentences within a time limit of 3 min.

Numerical processing

All numerical processing tasks were presented on notebooks running Presentation software. In all tasks, the background of the screen was black and the items were presented in white color in the middle of the screen. Participants were tested individually in a quiet room in school.

Dot enumeration. Sets of randomly arranged dots ranging from one to eight were presented which children had to enumerate as quickly as possible. The response was given by simultaneously pressing the space button and pronouncing the number. The experimenter recorded correctness. The key press initiated a mask (block pattern) for 1500 ms which prevented counting based on an after image of the dot display. The 48 trials (six per dot number) were presented in a fixed pseudo-random order

with the proviso that no dot number occurred twice in succession (interstimulus-interval: 1120 ms).

Single digit comparison. Children were presented with 56 pairs of digits and selected as quickly as possible the numerically larger one by pressing the corresponding keyboard button. Numerical distances ranged from 1 to 8 (16 trials for distance 1, ten trials for distance 2–3, and four trials each for distances 4–8). Stimuli were written in a 36-point Times New Roman font and presented in a randomized order, beginning with six practice trials (interstimulus-interval: 560 ms).

Magnitude comparison. Two gray displays with different numbers of yellow squares appeared side by side on the screen and children selected the numerically larger one as quickly as possible by keypress response (see **Figure 1**). Displays presented between 20 and 72 squares, and numerical distances between the two displays ranged from eight to 25 squares. Relatively high numerosities ensured that children based their decisions on estimation and not on verbal counting. The total surface areas in the two displays were identical. Each display consisted of different square sizes to avoid displays with larger numerosities systematically consisting of smaller squares. After three practice trials with feedback, 72 test trials (four for each numerical distance) were presented in a fixed pseudo-random order (interstimulus-interval: 300 ms).

Physical comparison. Here, children had to select the physically larger of two Arabic digits while ignoring their numerical value. In 32 trials, physical and numerical size were congruent (e.g., 2 6), in further 32 items physical and numerical size were incongruent (e.g., 2 6), and 18 neutral items displayed the same digit twice in different sizes (e.g., 2 2). Print sizes were a 48- and 24-point font. After six practice trials, the items were presented in random order (interstimulus-interval: 560 ms).

Comparison of two-digit numbers. Children were asked to select the numerically larger of two two-digit numbers between 21 and 98. In 30 items, both decade and unit digit were larger in one number (compatible items, e.g., 41 75), in further 30 items, the decade digit was larger in one and the unit digit was larger in the other number (incompatible items, e.g., 41 26). Overall distance and problem size were matched between the two compatibility conditions. All items had small numerical distance between the

FIGURE 1 | Example item of the magnitude comparison task (small distance).

decade digits and large distance between the unit digits (e.g., 37 52) as previous evidence (Nuerk et al., 2004) suggested that a compatibility effect was most likely to appear under these conditions. Twenty neutral items only differing in the unit digit (e.g., 61 68) were included in order to prevent children from basing their decisions on the decade digits only. Items were presented in a random sequence (interstimulus-interval: 560 ms).

Number line task. This task was adapted from Siegler and Opfer's (2003) number-to-position task. A number line (25 cm) was presented with the left end always labeled "0," and the right end labeled "100" for the first 24 items and "1000" for the next 24 items. Numbers in the lower range were overrepresented to allow discrimination between logarithmic and linear functions. An Arabic number appeared on top of the screen, and children read it out loud. Transcoding errors (which were exceptional) were corrected by the experimenter. Children indicated where the number would fall on the line by pointing with a cotton bud. The experimenter placed the cursor on this position and clicked the mouse. The deviance from the precise position was calculated in pixels (1 cm corresponded to 37.5 pixels). Each condition was introduced by three practice items. At t1 (beginning of Grade 2), only the number line 0–100 was given as the number line 0–1000 was considered too difficult.

RESULTS
ARITHMETIC PERFORMANCE
Table 1 presents for each of the three assessment points (end of Grade 1, beginning of Grades 3 and 4) the overall test scores for the two groups which constituted the selection criterion. **Table 1** also presents the raw scores for four of the arithmetic subtests, representing the number of simple calculation problems were answered correctly within two minutes. These subtest raw scores indicate a dramatic difference between the two groups: Dyscalculic children's scores were consistently more than 2 *SDs* below the number of items that were solved by children with age-adequate development.

NONVERBAL INTELLIGENCE, VERBAL WORKING MEMORY, ATTENTION, AND READING
Since low performance in any of these standardized tests was used as an exclusion criterion, performance of all participants was within average range. Still, average performance of the typically developing group was significantly better compared to the dyscalculia group on each of these measures.

NUMERICAL PROCESSING
Reliability was very high for all response time based measures (Cronbach's Alpha ranging between 0.93 and 0.98 for RTs at each assessment point) and sufficiently high for the two mental number line conditions (between 0.72 and 0.90, all *p*s < 0.01). Individual median reaction times were calculated for each child in each condition. Only reaction times for correct responses were considered, and reaction times lower than 200 ms and higher than 10000 ms were excluded. With the exception of the number line task where mean deviance scores were analysed, the main dependent variable used in statistical analysis was inverse efficiency (IE),
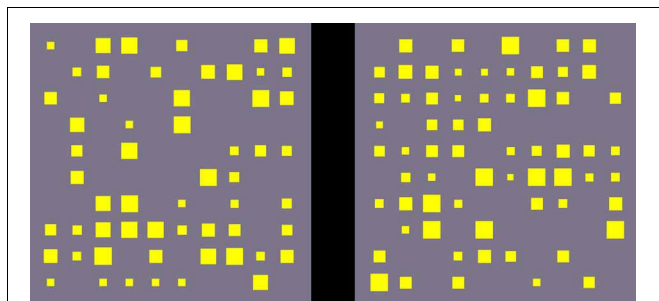
which combines accuracy and speed of response into one measure by dividing the adjusted median reaction times by the proportion of correct responses (Bruyer and Brysbaert, 2011).

Statistical analysis of each task was achieved by ANOVAs including all relevant within-task factors as well as assessment point (t1–t5) as within-subjects factors and arithmetic level (typical vs. dyscalculic) as between-subjects factor. In case of violation of sphericity, Greenhouse-Geisser correction was applied. Significant effects were followed up by paired comparisons under Bonferroni correction. Stability of task performance across the study period was examined by inspecting the correlational patterns between the five assessment points.

EFFICIENCY OF NUMERICAL PROCESSING

First, we wanted to know whether the efficiency to process numbers developed specifically or whether it was mainly dependent on increases in general processing speed. **Figure 2** presents for each assessment point children's IE-scores in the neutral condition of the physical comparison task with the average IE-scores in the digit comparison task. Although numbers are presented in the physical comparison task, the neutral condition requires a decision based on the physical size of two identical digits (e.g., 2 2) and therefore provides a non-numerical control measure of children's efficiency to perform forced-choice paradigms. Response accuracy was close to ceiling in both conditions even at t1 so that IE-scores mostly represented response times. Children showed systematically higher IE-scores in the numerical condition than in the physical condition, $F_{(1, 80)} = 458.65$, $p = <0.01$; $\eta^2 = 0.85$, IES-scores decreased systematically over time, $F_{(2.81, 224.57)} = 161.18$, $p < 0.01$ (all $ps < 0.05$) and dyscalculic children showed generally higher IE-scores than typically developing children, $F_{(1, 80)} = 23.29$, $p < 0.01$; $\eta^2 = 0.23$. **Figure 2** shows that IE-scores of dyscalculic and typically developing children were similar in the non-numerical condition (all $ps > 0.05$). In the numerical condition, however, dyscalculic children

FIGURE 2 | Efficiency of numerical processing: Mean IE-scores (SEs as error bars) for numerical (N) and physical (P) comparison across five assessment points.

showed clearly higher IE-scores than their typically developing peers, resulting in a significant task × arithmetic level interaction, $F_{(1, 80)} = 35.78$, $p < 0.01$, $\eta^2 = 0.31$. The interactions task × assessment point and task × assessment × arithmetic level were also reliable, $F_{(2.94, 244.57)} = 24.28$, $p < 0.01$; $\eta^2 = 0.23$, and $F_{(2.94, 234.85)} = 0.4.95$, $p < 0.05$.; $\eta^2 = 0.06$. *Post-hoc* analysis indicated that the difference in IE-scores between the two conditions decreased systematically over time among dyscalculic children (all $ps < 0.001$ except t1 vs. t2 and t4 vs. t5 where $p = 0.07$ and t3 vs. t4 and t5 where $p > 0.1$). The developmental change of these difference scores was smaller and not always significant among the typically developing children ($p < 0.05$ for t1 > t2, t4, t5, and t3 > t4, t5).

The stability of the efficiency of numerical processing across the study period was confirmed by mostly moderate correlations (ranging between 0.53 and 0.79, $p < 0.001$) among the difference scores (numerical minus physical condition) at each assessment point. The correlation between t1 and t4 appeared to be somewhat lower (0.23, $p = 0.03$) and the correlation between the two final assessment points (t4 and t5) was particularly high (0.84).

DOT ENUMERATION

For each child, the best fitting regression lines were calculated separately for the subitizing range (1–3) and the counting range (5–7)[2]. The regression lines are presented in **Figure 3**.

Because of very obvious and expected differences between IE-scores for subitizing and counting, intercepts and slopes for these two numerical ranges were analysed separately. In the subitizing range, the main effect of assessment point was significant for intercepts, $F_{(3.07, 242.24)} = 25.40$, $p < 0.01$; $\eta^2 = 0.243$, while slopes remained largely constant across the study period, $F_{(3.07, 242.11)} = 0.52$, n.s.; $\eta^2 = 0.007$. Interestingly, intercepts as well as slopes turned out to be larger in the dyscalculic than in the typically developing group, intercepts: $F_{(1, 79)} = 10.55$, $p = 0.002$; $\eta^2 = 0.12$, slopes: $F_{(1, 79)} = 10.36$, $p < 0.01$; $\eta^2 = 0.12$. The interaction approached significance for intercepts, $F_{(3.07, 242.24)} = 2.40$, $p = 0.067$; $\eta^2 = 0.030$, but not for slopes, $F_{(3.06, 242.11)} = 0.98$, n.s.; $\eta^2 = 0.012$. *Post-hoc* analysis indicated larger subitizing intercepts at t1 than at all later assessment points for the dyscalculia group while the decrease in intercepts over time was more systematic in the typically developing group (t1 > t2, t4, t5; t2 > t4, t5, t3 > t5).

In the counting range, the main effect of assessment point was again significant for intercepts, $F_{(2.27, 179.33)} = 23.31$, $p < 0.05$.; $\eta^2 = 0.228$ (*post-hoc* tests: t1 < t2, t3 < t4, t5, all $ps < 0.06$), but not for slopes, $F_{(1.68, 132.22)} = 2.11$, n.s.; $\eta^2 = 0.026$. Intercepts were significantly larger in the dyscalculia than in the typically developing group, $F_{(1, 79)} = 34.05$, $p < 0.05$.; $\eta^2 = 0.301$. The effect of arithmetic level was only of borderline significance for the counting slopes: $F_{(1, 79)} = 3.86$, $p = 0.053$; $\eta^2 = 0.047$. No
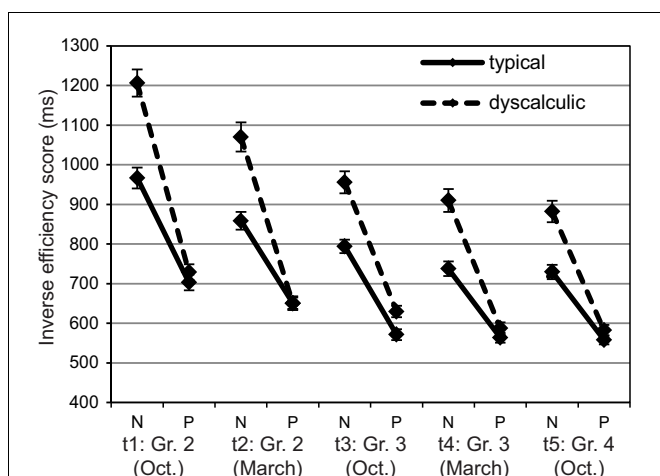
---

[2]IE-scores for eight dots were excluded because previous studies (e.g., Trick and Pylyshyn, 1993) demonstrated end effects for the enumeration of the largest numerosity of a set. IE-scores for four dots were also excluded, as earlier studies (Schleifer and Landerl, 2011; Reeve et al., 2012) suggested that in this intermediate range, some children might still be able to subitize, while others would already resort to their counting skills.

assessment point × arithmetic level interactions were evident for the counting range ($Fs = 1.12$ and 0.33, n.s.).

Low to medium range correlations were observed among the intercepts of the first four assessment points for both, subitizing (0.28, $p = 0.011$ to 0.67, $p < 0.001$) and counting (0.32, $p = 0.003$ to 0.64, $p < 0.001$). Stability was generally low for subitizing slopes across the first four assessment points, with correlations in the moderate range (0.22–0.34, $ps < 0.05$, except t1 and t3 with $r = 0.15$, n.s.). Counting slopes showed some stability between t1 and t2 (0.28, $p = 0.011$), while no significant correlations were evident between t2, t3, and t4. Higher stability for the dot counting task was found between t4 and t5, with high correlations between the intercepts for subitizing (0.78) and counting (0.80), as well as between the counting slopes

(0.90, all $ps < 0.001$). The correlation between the subitizing slopes at t4 and t5 was in the medium range with $r = 0.38$, $p < 0.001$.

SINGLE DIGIT COMPARISON[3]

In order to display the effect of numerical distance, the best fitting regression line was calculated for each child. For intercepts, both main effects were significant, assessment point: $F_{(2.94, 235.41)} = 87.47$, $p < 0.01$; $\eta^2 = 0.55$; arithmetic level: $F_{(1, 80)} = 37.00$, $p < 0.01$; $\eta^2 = 0.82$. **Figure 4** shows a continuous decrease of

---

[3]Due to unrealistic scores at t3, one dyscalculic child had to be excluded from analysis of the digit comparison task.
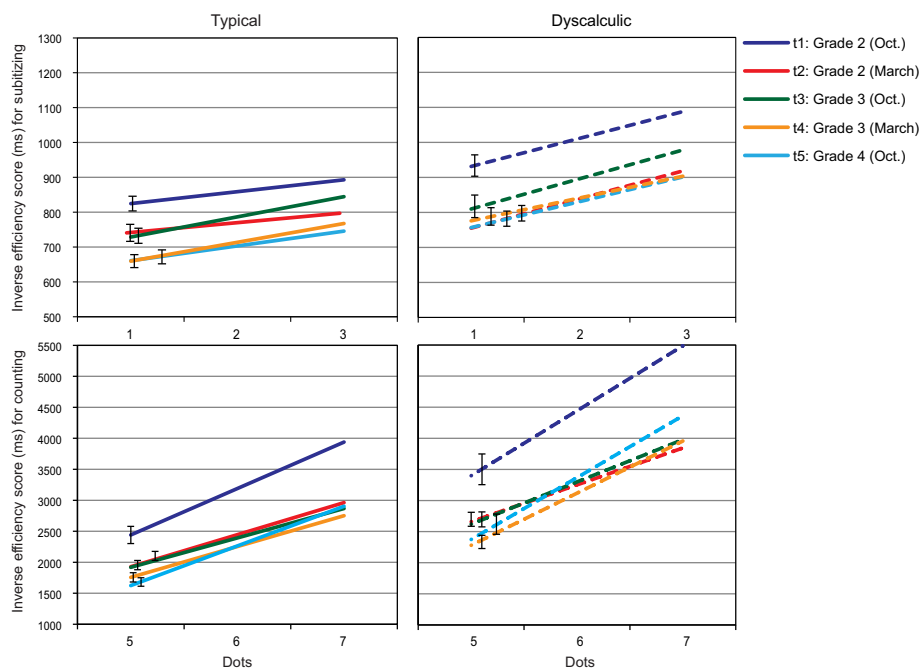
FIGURE 3 | Dot enumeration: Regression lines for IE-scores in the subitizing range (top) and the counting range (bottom) across five assessment points (error bars represent SEs of intercepts).
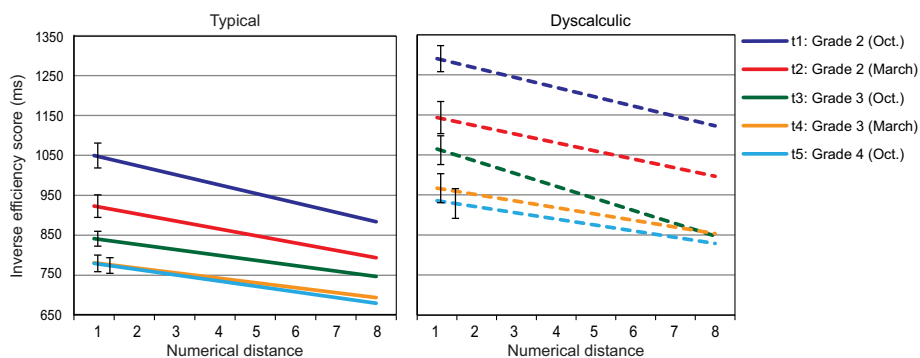
FIGURE 4 | Single digit comparison: regression lines of IE-scores in relation to numerical distance across five assessment points (error bars represent SEs of intercepts).

intercepts from t1 to t5 (t1, t2 > t3, t4, and t3 > t5; $ps < 0.05$). Most importantly, intercepts of the dyscalculia group were consistently higher than those of the control group at each assessment point (all $ps < 0.01$).

For slopes, the main effect of assessment point was significant, $F_{(2.96, 236.51)} = 3.48$, $p < 0.05$; $\eta^2 = 0.042$, and the interaction assessment point $\times$ arithmetic level approached significance: $F_{(2.96, 236.51)} = 2.35$, $p = 0.074$; $\eta^2 = 0.029$. Visual inspection of **Figure 4** as well as *post-hoc* analysis indicated that this interaction was driven by a group difference at t3 ($p < 0.001$), but not at the other assessment points (all $ps > 0.05$).

Thus, although the overall efficiency of numerical processing (represented by intercepts) showed a clear improvement across the study period and a considerable developmental delay for the dyscalculia group, we could not confirm earlier evidence from cross-sectional studies (Holloway and Ansari, 2008; Landerl and Kölle, 2009) reporting a decrease in numerical distance effect over time. One could even argue that there was a relative increase of the distance effect when taking the decrease in intercepts into consideration. However, when this was examined in an additional ANOVA where slopes divided by intercepts were subjected as dependent measures, no significant effects remained.

Robust correlations ranging between 0.62 and 0.90 (all $ps < 0.001$) were observed between the intercepts of the five assessment points. For slopes, however, a reasonable amount of stability was only evident from t3 on (0.36, 0.52, and 0.64 for correlations between t3 and t4, t3 and t5 and t4 and t5, respectively, all $ps < 0.001$), while no significant correlations were found with the earlier assessment points.

MAGNITUDE COMPARISON[4]

In order to investigate the non-symbolic distance effect the items of this task were grouped into two distance levels: (1) small distance condition (differences between the two displays ranged from 8 to 16) and (2) large distance condition (differences between the two displays ranged from 17 to 25). From **Figure 5**, it is obvious that all groups were faster in responding to large distance items than to small distance items. In a 2 (distance) $\times$ 5 (assessment point) $\times$ 2 (arithmetic level) ANOVA, the

---

[4]As the task design was not based on ratios between the two displays, calculation of Weber fractions was not possible.

main effect of distance, $F_{(1, 81)} = 477.04$, $p < 0.01$; $\eta^2 = 0.86$, was indeed highly reliable. The main effects of assessment point, $F_{(2.22, 179.68)} = 81.97$, $p < 0.01$; $\eta^2 = 0.50$ was modified by a significant interaction distance $\times$ assessment point, $F_{(3.01, 243.82)} = 19.58$, $p < 0.01$; $\eta^2 = 0.20$. *Post-hoc* analysis indicated significant differences between all assessment points for small as well as large distance items ($ps < 0.01$). The interaction was caused by a relatively small decrease in IE-scores from t4 to t5 for items with a large numerical distance.

Importantly, there was an effect of arithmetic level, $F_{(1, 81)} = 11.17$, $p = 0.03$; $\eta^2 = 0.12$, which did not interact with the other factors, indicating that the dyscalculia group showed lower performance over all, while the pattern of performance was comparable throughout the study period. This was confirmed in a final ANOVA calculating a relative distance effect as percent increase of IE-scores in the small compared to the large distance condition. In this analysis, only the main effect of assessment point remained significant, $F_{(3.31, 268.10)} = 8.63$, $p < 0.01$; $\eta^2 = 0.10$.

There were medium-sized correlations among IE-scores for small as well as large distance items between t1 and t4 (ranging from 0.38 to 0.56, all $ps < 0.001$) and high correlations between t4 and t5 (0.95 for both, small and large distance items). For the non-symbolic distance effect itself, only moderate correlations were observed between t1 and the later assessment points (rs between 0.20, $p = 0.07$, and 0.38, $p < 0.001$) and t3 and the later assessment points ($r = 0.21$, $p = 0.06$ for t4 and 0.30, $p = 0.007$ for t5). High stability for the non-symbolic distance effect was only achieved between the final two assessment points ($r = 0.76$, $p < 0.001$).

PHYSICAL COMPARISON

Figure 6 shows a very systematic size congruity effect for both typically developing and dyscalculic children. In a 3 (congruity) $\times$ 5 (assessment point) $\times$ 2 (arithmetic level) ANOVA we found significant main effects of congruity, $F_{(1.96, 158.43)} = 34.91$, $p < 0.01$; $\eta^2 = 0.30$, and assessment point, $F_{(2.77, 224.53)} = 79.11$, $p < 0.01$; $\eta^2 = 0.48$, but no difference between groups and no interactions involving group. *Post-hoc* analyses showed significantly lower IE-scores for congruent than for neutral items (facilitation effect), and again lower IE-scores for neutral than for incongruent items (interference effect) ($ps < 0.01$). IE-scores decreased
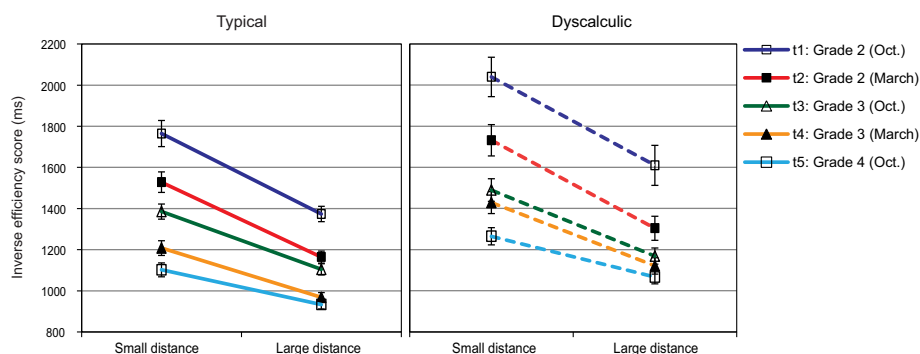
FIGURE 5 | Magnitude comparison: Mean IE-scores (SEs as error bars) for small and large numerical distance items at each assessment point.
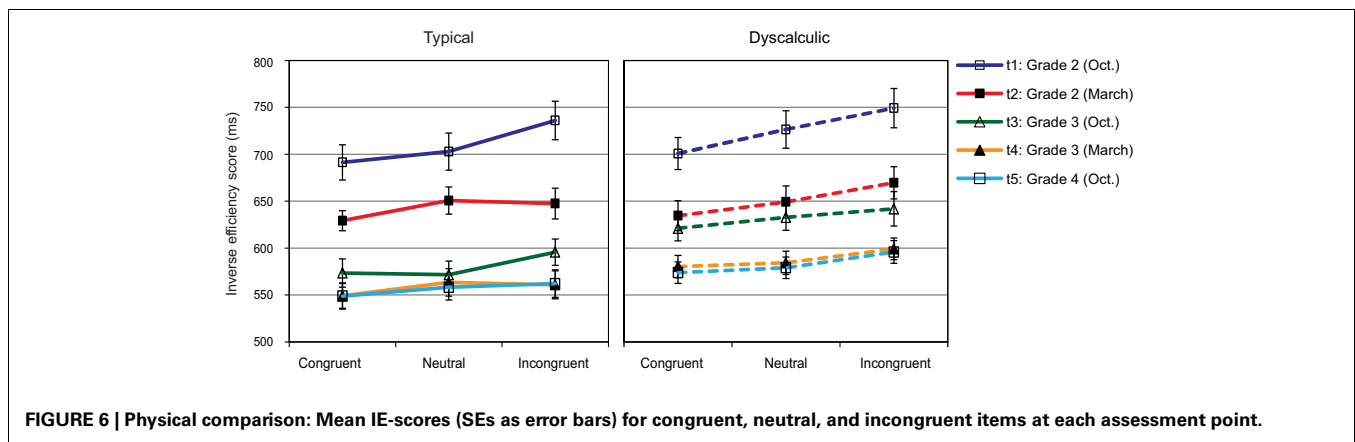
FIGURE 6 | Physical comparison: Mean IE-scores (SEs as error bars) for congruent, neutral, and incongruent items at each assessment point.

systematically from t1 to t5 (all $ps < 0.05$). The only significant interaction was found between congruity and assessment point: $F_{(5.61, 454.45)} = 2.23$, $p = 0.043$; $\eta^2 = 0.03$. Pairwise contrasts indicated a significant difference between t3 and t4 in the congruent and the incongruent, but not in the neutral condition.

Thus, the interesting finding from this task was that even at the earliest assessment point at the beginning of Grade 2, typically developing and even dyscalculic children showed a significant influence of the numerical value of two presented digits on their non-numerical decision. In contrast to our expectation that the size of the facilitation and interference effects should increase with experience, these effects did not change much across the whole period of the study, in spite of a general incease in processing efficiency. In a final analysis, it was examined whether there were relative differences in the facilitaion and interference effect when these effects were expressed as % change of IE-scores in relation to the neutral condition. Mostly because variability between participants was high, none of the effects remained significant.

Although there were moderate correlations among the first four assessment points for IES-scores in each of the three conditions (rs between 0.36 and 0.55, all $ps < 0.001$) and high correlations between t4 and t5 (congruent: 0.96, neutral: 0.95, incongruent: 0.96), the only correlations over time for the facilitation and the interference effect became evident between t4 and t5 (0.71 and 0.73, $p < 0.001$).

COMPARISON OF TWO-DIGIT NUMBERS

This was the only task in our numerical processing battery for which response accuracy was not close to ceiling and therefore had a considerable impact on IE-scores. Especially at t1, both groups showed considerable problems with the incompatible condition with 73.73% correct for typically developing and only 50.57% correct for the dyscalculia group. Typically developing children showed mean response accuracies above 90% for later assessments of incompatible items and for compatible items throughout. The dyscalculia group reached this high level of performance only at t4 for incompatible items (88.20% correct) and even in the compatible condition, only 87.37% of the items were responded to correctly at t1. IE-scores, which integrate these accuracy scores with children's speed of response, are plotted in **Figure 7**.

In a 2 (compatibility) × 5 (assessment point) × 2 (arithmetic level) ANOVA, all main effects were significant: compatibility, $F_{(1, 81)} = 56.42$, $p < 0.01$; $\eta^2 = 0.41$, assessment point, $F_{(1.19, 96.20)} = 59.15$, $p < 0.01$; $\eta^2 = 0.42$, arithmetic level, $F_{(1, 81)} = 32.08$, $p < 0.01$; $\eta^2 = 0.28$. Overall, children had higher IE-scores in the incompatible than in the compatible condition. IE-scores decreased systematically during the study period (t1 > t2, t3 > t4 > t5, all $ps < 0.01$), and children with dyscalculia had higher IE-scores than typically developing children. In addition, all interactions were reliable, compatibility × assessment point: $F_{(1.15, 92.76)} = 24.19$, $p < 0.01$; $\eta^2 = 0.23$, compatibility × arithmetic level: $F_{(1, 81)} = 16.17$, $p < 0.01$; $\eta^2 = 0.17$, assessment point × arithmetic level: $F_{(1.19, 96.20)} = 5.70$, $p < 0.05$; $\eta^2 = 0.07$, compatibility × assessment point × arithmetic level: $F_{(1.145, 92.76)} = 6.88$, $p < 0.01$; $\eta^2 = 0.78$.

In order to interpret this complex pattern of interactions and to analyze the particular problems children face when processing two-digit numbers, two additional ANOVAs were calculated: First, we subjected the IE-scores for the easier condition of compatible items to a 5 (assessment point) × 2 (arithmetic level) ANOVA. Both main effects were reliable, assessment point: $F_{(2.18, 176.20)} = 61.38$, $p < 0.01$; $\eta^2 = 0.43$, arithmetic level: $F_{(1, 81)} = 21.68$, $p < 0.01$; $\eta^2 = 0.21$, and the two factors also interacted, $F_{(2.18, 176.20)} = 4.23$, $p < 0.05$; $\eta^2 = 0.05$. *Post-hoc* analysis indicated a systematic decrease of IE-scores during the study period for the dyscalculic children (all $ps = 0.001$ except t1 vs. t2), while among typically developing children the IE-score differences between adjacent assessment points were too small to be reliable from t2 on (all $ps < 0.06$ except t2 vs. t3, t3 vs. t4 and t4 vs. t5). The difference between the two groups was significant at all assessment points except t1, which is obviously due to the high variability (especially among the dyscalculic children).

The developmental trajectories of the compatibility effect were analysed by subtracting IE-scores for compatible items from those for incompatible items. This difference score was again subjected to a 5 (assessment point) × 2 (arithmetic level) ANOVA. Again, both main effects and the interaction were reliable, assessment point: $F_{(1.15, 92.76)} = 24.19$, $p < 0.01$; $\eta^2 = 0.21$, arithmetic level: $F_{(1, 81)} = 16.17$, $p < 0.01$; $\eta^2 = 0.17$, assessment point × arithmetic level: $F_{(1.15, 92.76)} = 6.88$, $p < 0.01$; $\eta^2 = 0.08$. As evident from **Figure 7**, both groups showed an especially strong
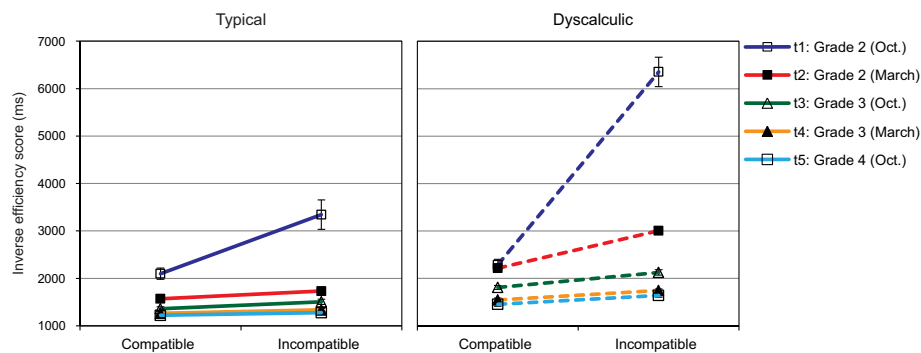
FIGURE 7 | Comparison of two digit numbers: Mean IE-scores (SEs as error bars) for compatible and incompatible items across five assessment points.

compatibility effect at t1, but this effect was still significantly larger ($p < 0.01$) for dyscalculic (4067) than for typically developing children (1246 ms). Among the typically developing children the compatibility effect was relatively small and similar for the later assessment points (t2: 163 ms; t3: 144 ms: t4: 71 ms, t5: 55 ms). In the dyscalculia group, however, particular problems to integrate the tens and units of two-digit numbers were still evident at t2: Their compatibility effect (789 ms) was significantly larger compared to typically developing children at t2 and larger than at the later assessment points (t3: 316 ms, $p = 0.07$; t4: 200 ms, $p = 0.019$; t5: 185 ms, $p = 0.013$). Group differences between typically developing and dyscalculic children were still marked at the later assessment points, t3: $p = 0.08$; t4 and t5: $p < 0.05$.

In summary, these longitudinal data clearly showed that efficient processing of two-digit numbers develops slowly and poses a particular challenge to children with dyscalculia. Correlations of IE-scores for incompatible items were only moderate for t1 with later assessment points (between 0.32 and 0.34, $p = 0.002$) and in the medium range for t2, t3, and t4 as well as for compatible items among the first four assessment points (rs between 0.48 and 0.77). Correlations were considerably higher between t4 and t5 (0.95 and 0.91 for compatible and incompatible items, respectively). For the compatibility effect itself, moderate correlations were found between t2 and t3 ($r = 0.37$, $p = 0.001$), t3 and t4 ($r = 0.24$, $p = 0.026$). Once again, reasonable stability was only evident between t4 and t5, $r = 0.73$, $p < 0.001$.

NUMBER LINE TASK

Two separate ANOVAs with median deviance in pixel (see **Figure 8**) as dependent variable were calculated for number lines 0–100 and 0–1000. For number line 0–100, assessment point as well as arithmetic level showed reliable effects, $F_{(2.18, 176.16)} = 121.93$, $p < 0.01$; $\eta^2 = 0.60$, and $F_{(1, 81)} = 53.37$, $p < 0.01$; $\eta^2 = 0.38$, which were modulated by a significant interaction, $F_{(2.18, 176.16)} = 29.03$, $p < 0.01$; $\eta^2 = 0.26$. Children with dyscalculia showed higher deviance scores than their typically developing peers at all assessment points (all $ps < 0.05$), but this group difference decreased across assessment points (effect sizes of 0.66,
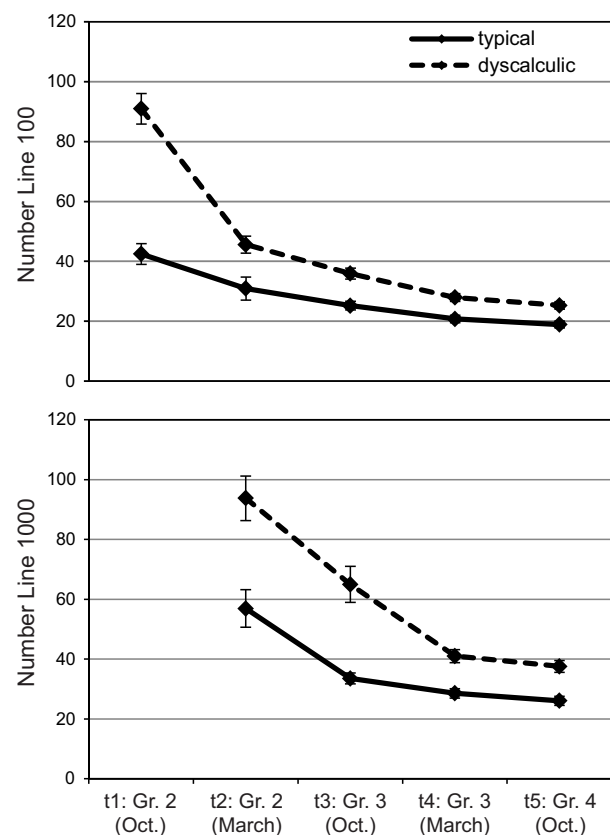
FIGURE 8 | Number line: Mean deviance scores in pixel (SEs as error bars) across five assessment points for number lines 100 (top) and 1000 (bottom).

0.32, 0.46, 0.42, and 0.40 for t1–t5, respectively). Typically developing children's performance improved significantly ($ps < 0.01$) between t1 and t3/t4/t5, t2 and t4/t5, as well as t3/t4 and t5. Dyscalculic children showed significant improvements between all assessment points.

The 0–1000 number line condition (lower section of **Figure 8**) was not given at t1 as it was assumed to be too difficult for

children at the beginning of Grade 2. Once again, we found significant main effects for assessment point, $F_{(1.78, 140.89)} = 61.17$, $p < 0.01$; $\eta^2 = 0.43$ and arithmetic level as well as an interaction, $F_{(1, 81)} = 30.94$, $p < 0.01$; $\eta^2 = 0.28$. Children with dyscalculia showed significantly larger deviances from the correct response than children with typical arithmetic development at all assessment points ($ps < 0.01$). The interaction resulted from the relatively high variability in t1 compared to later assessment points. Among typically developing children, significant improvements were observed between t2 and t5 as well as between all later assessment points. Dyscalculic children showed significant improvements between all four assessment points (all $ps < 0.01$).

In order to test earlier claims (Siegler and Booth, 2004) that children's mental number line progresses from a logarithmic to a linear representation over time, we calculated both regression lines separately for each child and each assessment point. No such developmental change could be observed for the typically developing children for whom a linear fit (R^2s between 0.88 and 0.99) was found to describe children's performance better than a logarithmic fit (R^2s between 0.69 and 0.79) at all assessment points and in both conditions (all $ps < 0.01$ in Wilcoxon signed-rank tests). For the dyscalculia group, a logarithmic fit ($R^2 = 0.71$) seemed somewhat more adequate than a linear fit ($R^2 = 0.64$, $p < 0.01$) at t1 where only the number line 0–100 was given, however, both fits were rather low. At all later assessment points, a linear fit (R^2s between 0.89 and 0.99) described dyscalculic children's performance clearly better than a logarithmic fit (R^2s between 0.70 and 0.76, all $ps < 0.01$).

Correlations between assessment points were moderate for both conditions among the first four assessment points (rs between 0.33 and 0.62, all $ps < 0.005$) and clearly higher between t4 and t5 (0.97 and 0.99).

DISCUSSION

In this study, typically developing and dyscalculic children's development of numerical processing was followed from the beginning of Grade 2 until the beginning of Grade 4. While most earlier studies investigating basic numerical processing in elementary school either compared different age groups cross-sectionally or covered relatively short periods of development, the present design allowed us to examine intra- as well as interindividual differences in numerical processing during a 2-year period which constitutes an important phase of arithmetic development. In these elementary school years, the foundations of arithmetic like place-value system, mental arithmetic, and written calculations are taught and practiced in school. Numerical processing has been demonstrated to be associated with these arithmetic skills (Halberda and Feigenson, 2008; de Smedt et al., 2009; Jordan et al., 2009; Geary, 2011), which is why it seemed particularly interesting to focus on this developmental period. Another crucial reason for the focus on these early school years was that this is when problems in arithmetic development become obvious and dyscalculia is diagnosed. For the present purpose, we selected groups of children who showed either age-adequate or persistently poor arithmetic performance during the study period. Marked and persistent problems in arithmetic in spite of adequate general cognitive abilities are the central diagnostic

criterion of dyscalculia (e.g., World Health Organization, 2010). It is as yet unclear which subcomponents of numerical processing are central to arithmetic development, therefore we decided to apply a battery of tasks that have been used before to assess standard effects of symbolic as well as non-symbolic numerical processing.

INDICATORS OF NUMERICAL DEVELOPMENT DURING ELEMENTARY SCHOOL

A first finding was that all investigated effects of numerical processing were evident as early as Grade 2 for both typically developing children and children with persistent arithmetic problems. This is mostly consistent with earlier studies (e.g., Sekuler and Mierkiewicz, 1977; Girelli et al., 2000; Landerl and Kölle, 2009; Pixner et al., 2009; Reeve et al., 2012). It was, however, surprising for the physical comparison task. The size-congruity effect requires a certain amount of experience-based automaticity in numerical processing. It is influenced by certain task characteristics (most importantly the difference in physical size between the two presented digits, see Schwarz and Ischebeck, 2003) and earlier studies already indicated that there is a good deal of variability in when it appears. Importantly, in the current study, not only typically developing but also dyscalculic children showed sufficient automaticity in numerical processing to produce facilitation and interference effects even at the beginning of Grade 2. Together with the finding that there was no increase of these effects over time we confirmed Bugden and Ansari's (2011) claim that automatic processing of Arabic numerals is not directly related to arithmetic skills.

Generally, efficiency of numerical processing turned out to be a very good indicator of numerical development. Throughout the study period, we observed a systematic increase in speed of processing for numbers, which was larger than the general increase in processing speed that is characteristic for child development (Kail, 1991). Furthermore, the dyscalculia group showed persistent deficiencies in the speed of processing which were specific to numerical information and did not extend to non-numerical comparisons.

While efficiency of numerical processing improved consistently, many of the investigated within-task effects of numerical processing remained largely constant across time, for both typically developing and dyscalculic children. More specifically, while some earlier cross-sectional studies had suggested that the distance effect would decrease over time indicating an incremental specification of the cognitive representation of number (Holloway and Ansari, 2008; Landerl and Kölle, 2009), our longitudinal data showed no such decrease, neither for the symbolic nor for the non-symbolic comparison task. It might be argued that in relation to the decreasing intercepts, slopes that remain constant across time in fact indicate a relative increase of the investigated effect. In other words, when overall numerical processing becomes more efficient during development, it could be expected that within-task effects should decrease in accordance with intercepts, while our evidence suggests that slopes did not change much. However, even when within-task effects (slopes) were expressed as changes of IE-scores relative to overall efficiency of numerical processing (intercepts), no significant differences

appeared across assessment points or arithmetic level, which confirms that the symbolic and non-symbolic distance effects did not undergo marked changes during the study period. The most likely explanation for this negative evidence is that within-task effects were relatively small while intra- as well as interindividual variability of task performance was relatively high. Correspondingly, stability was found to be low until middle of Grade 3.

STABILITY OF NUMERICAL PROCESSING

Although we found moderate to medium-range correlations across assessment points for intercepts, correlations for experimental effects were mostly non-significant. Only between the last two assessment points of our study period, i.e., middle of Grade 3 and beginning of Grade 4, robust correlations were evident for all measures. In summary, significant long-term stability could be observed for the overall efficiency of numerical processing. Yet, reasonable stability of within-task effects of numerical cognition was only achieved toward the end of primary school, but was found to be low in the early phases. Note that the 83 participants of the present study were specifically selected because their arithmetic competence showed a relatively steady development over time. It is likely that stability of numerical processing is even lower in the full sample of 139 children (including those who showed a variable arithmetic profile) and in the general population. Reeve et al. (2012) have recently reported reasonable stability for a random sample across a 6-year period for dot enumeration and symbolic number comparison. This analysis was also mostly based on speed of response and limited to ordinal correlations of group membership. Reeve et al.'s finding that 69% of the sample remained in the identified "slow," "medium," and "fast" groups implicates that almost one third of the participants exhibited considerable variability in their numerical processing development.

DYSCALCULIA

A main research question of the current study was whether the numerical development of children with persisting arithmetic problems is mostly delayed or whether it would be possible to identify dyscalculia-specific anomalies in numerical cognition. As already mentioned, the dyscalculic children showed serious and pervasive deficiencies with respect to efficiency of numerical processing. Importantly, these problems were not limited to those tasks that required processing of symbolic representations of number, but were also evident for the magnitude comparison task. Thus, the current data do not provide support for Rousselle and Noël's (2007) proposal of specific problems to access numerical information from symbolic representations. However, note again that although the dyscalculia sample performed at a systematic lower level in the comparison paradigms, they showed symbolic as well as non-symbolic distance effects that were not significantly different from the typically developing children. Thus, we could not confirm Mussolin et al.'s (2010) finding of stronger symbolic and non-symbolic distance effects in dyscalculia. It is possible that problems to differentiate between two numbers or quantities are more prominent for smaller distances due to higher numerical similarity. This might explain as to why the Mussolin et al. study, which examined distances up

to only four, found a reliable difference that we did not detect. Furthermore, their sample was somewhat older (10–11 years) and smaller and may have performed more homogeneously than the current sample. We can also not rule out that a ratio-based design of magnitude comparison might have revealed lower acuity of the approximate number system as it was reported before (see Piazza et al., 2010). Based on the findings of the current repeated assessment of the distance effect, we conclude that although dyscalculic children have marked problems to access their numerical cognition system efficiently, we did not find evidence for abnormal cognitive representations of numerosities in symbolic and non-symbolic comparison paradigms.

Anomalies were, however, evident in the dot enumeration paradigm where the dyscalculia group showed not only larger intercepts, but also persistently larger slopes in the subitizing range. For the higher numbers of the counting range, group differences in slopes were less marked. This evidence provides further support for earlier claims of a particular subitizing problem in dyscalculia (Moeller et al., 2011; Schleifer and Landerl, 2011; Reeve et al., 2012). Butterworth (2010) has argued that subitizing may reflect an inborn capacity to quantify over sets which provides the foundation for associating numbers with distinct numerosities. Such an early deficit may well induce problems in mapping between numbers and quantities and in the long run a general inefficiency in numerical processing as it was observed in the current data set. Over time, it would also induce a general imprecision of numerical representations. This is exactly what we found in the number line task: Over the whole study period, dyscalculic children showed larger deviances from the precise location of a number on a number line than children with typical development of arithmetic skills. The number line task is particularly important in the current design as it was the only untimed task in our numerical processing battery. The persistent deficit in the dyscalculia sample shows that their numerical processing problems are not limited to processing speed. Interestingly, earlier claims of a developmental trajectory from an overrepresentation of small numbers in the mental number line inducing a logarithmic function to a linear representation (e.g., Booth and Siegler, 2006), did not find support in the current data set (see also Landerl et al., 2009).

An important aspect that has not yet been thoroughly investigated in dyscalculia is the acquisition of the place-value system of the Arabic notational system. Our findings on processing of two-digit numbers add to current evidence on young children's difficulties to integrate ten and unit numbers (Nuerk et al., 2004; Pixner et al., 2009; Mann et al., 2011, 2012). In accordance with Pixner et al. (2009) we found particularly poor performance at the beginning of Grade 2, but rapid improvement in the competence to process two-digit numbers for the typically developing group. Dyscalculic children's problems were clearly more pronounced and persistent throughout the study period. At t1 they actually chose the incorrect, smaller number in about half of the trials. One might assume that they attempted to select the larger number based on the unit number which would induce systematically wrong choices in the incompatible condition. However, the fact that they chose the larger unit digit in about half of the items in the incompatible condition (and therefore responded correctly)

and even in about 15% of the compatible items (and therefore responded incorrectly) speaks against such a strategy and rather suggests that they were guessing. Note that in the German test-language, tens and units are inversed in numberwords (21 is "one and twenty") which has been demonstrated to amplify children's problems to acquire the place-value system (Pixner et al., 2011).

IMPLICATIONS FOR FUTURE RESEARCH AND DYSCALCULIA DIAGNOSIS

In summary, the current longitudinal data set shows that efficiency of numerical processing is an important indicator of numerical skills: Despite considerable improvements during the elementary school years it remains persistently deficient in children with dyscalculia. While significant stability was found for speed, many of the investigated within-task effects were of low stability and not subject to developmental processes. Because of the low stability of these effects across time, they do not seem appropriate for diagnostic tests of dyscalculia. The most obvious criterion to identify children who struggle with their numerical processing system is the efficiency of numerical processing. It will be important to devise more computerized tests enabling accurate measurement of response times as this is the main indicator of efficiency of numerical processing in simpler tasks like dot enumeration of number comparison.

While the finding of a generally lower efficiency of numerical processing suggests a delayed rather than a deviant numerical development in dyscalculia, the current study also helped to identify parameters that go beyond the developmental delay perspective: The dyscalculia sample showed persistently larger slopes in the subitizing range of dot enumeration, inaccurate numerical estimation in the number line task and serious problems to integrate the component numerals in multi-digit numbers. Subitizing seems to have a strong biological basis (Vetter et al., 2011) and may be a very early indicator of a faulty numerical processing system, while both, the number line task and processing of multi-digit numbers, develop as a consequence of education and experience. While the focus of the current study was the development of numerical processing in elementary school children who already experience persistent problems in arithmetic, future studies should concentrate on earlier phases of development in order to identify the developmental trajectories of the relevant parameters even before the problems in arithmetic arise.

ACKNOWLEDGMENTS

REFERENCES

Barth, H. C., and Paladino, A. M. (2011). The development of numerical estimation: evidence against a representational shift. *Dev. Sci.* 14, 125–135. doi: 10.1111/j.1467-7687.2010.00962.x

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., and Zorzi, M. (2010). Numerical estimation in preschoolers. *Dev. Psychol.* 46, 545–551. doi: 10.1037/a0017887

Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 42, 189–201. doi: 10.1037/0012-1649.41.6.189

Bruyer, R., and Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the Inverse Efficiency Score (IES) a better dependent variable than the mean Reaction Time (RT) and the Percentage of Errors (PE)? *Psychol. Belg.* 51, 5–13.

Bugden, S., and Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition* 118, 32–44. doi: 10.1016/j.cognition.2010.09.005

Butterworth, B. (1999). *The Mathematical Brain*. London: MacMillan.

Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends Cogn. Sci.* 14, 534–541. doi: 10.1016/j.tics.2010.09.007

Butterworth, B., Varma, S., and Laurillard, D. (2011). Dyscalculia: from brain to education. *Science* 332, 1049–1053. doi: 10.1126/science.1201536

Cattell, R. B. (1950). *Culture Fair (or free) Intelligence Test (A Measure of "g") Scale 1. Handbook for the Individual or Groups*. Champaign, IL: IPAT.

Cattell, R. B., Weiß, R. H., and Osterland, J. (1997). *Grundintelligenztest Skala 1 (CFT 1)*. Göttingen: Hogrefe.

de Smedt, B., Verschaffel, L., and Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *J. Exp. Child Psychol.* 103, 469–479. doi: 10.1016/j.jecp.2009.01.010.

Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics (Rev. ed.)*. New York, NY: Oxford University Press.

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., and Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year old children: Evidence for a segmented linear model. *J. Exp. Child Psychol.* 99, 1–17. doi: 10.1016/j.jecp.2007.08.006

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Dev. Psychol.* 47, 1539–1552. doi: 10.1037/a0025510

Girelli, L., Lucangeli, D., and Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *J. Exp. Child Psychol.* 76, 104–122. doi: 10.1006/jecp.2000.2564

Haffner, J., Baro, K., Parzer, P., and Resch, F. (2005). *Heidelberger Rechentest (HRT 1-4)*. Göttingen: Hogrefe.

Halberda, J., and Feigenson, L. (2008). Developmental change in the acuity of the "number sense": the approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Dev. Psychol.* 44, 1457–1465. doi: 10.1037/a0012682

Holloway, I. D., and Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. *Dev. Sci.* 11, 644–649. doi: 10.1111/j.1467-7687.2008.00712.x

Iuculano, T., Tang, J., Hall, C. W., and Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Dev. Sci.* 11, 669–680. doi: 10.1111/j.1467-7687.2008.00716.x

Jordan, N. C., Glutting, J., and Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learn. Individ. Diff.* 20, 82–88. doi: 10.1016/j.lindif.2009.07.004

Jordan, N. C., Kaplan, D., Nabors Olah, L., and Locuniak, M. N. (2006). Number sense growth in kindergarten: a longitudinal investigation of children at risk for mathematics difficulties. *Child Dev.* 77, 153–175. doi: 10.1111/j.1467-8624.2006.00862.x

Jordan, N. C., Kaplan, D., Ramineni, C., and Locuniak, M. N. (2009). Early math matters: Kindergarten number competence and later mathematics outcomes. *Dev. Psychol.* 45, 850–867. doi: 10.1037/a0014939

Kail, R. (1991). Developmental changes in speed of processing during childhood and adolescence. *Psychol. Bull.* 109, 490–501. doi: 10.1037/0033-2909.109.3.490

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., et al. (2011). Mental number line training in children with developmental dyscalculia.

Neuroimage 57, 782–795. doi: 10.1016/j.neuroimage.2011.01.070

Kucian, K., Loenneker, T., Martin, E., and von Aster, M. (2011). Non-symbolic numerical distance effect in children with and without developmental dyscalculia: a parametric fMRI study. *Dev. Neuropsychol.* 36, 741–762. doi: 10.1080/87565641.2010.549867

Landerl, K., Bevan, A., and Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8–9-year-old students. *Cognition* 93, 99–125. doi: 10.1016/j.cognition.2003.11.004

Landerl, K., Fussenegger, B., Moll, K., and Willburger, E. (2009). Dyslexia and dyscalculia: Two learning disorders with different cognitive profiles. *J. Exp. Child Psychol.* 103, 309–324. doi: 10.1016/j.jecp.2009.03.006

Landerl, K., and Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. *J. Exp. Child Psychol.* 103, 546–565. doi: 10.1016/j.jecp.2008.12.006

Libertus, M. E., Feigenson, L., and Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Dev. Sci.* 14, 1292–1300. doi: 10.1111/j.1467-7687.2011.01080.x

Mann, A., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2011). Attentional strategies in place-value integration. *Z. Psychol.* 219, 42–49. doi: 10.1027/2151-2604/a000045

Mann, A., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2012). On the development of Arabic three-digit number processing in primary school children. *J. Exp. Child Psychol.* 113, 594–601. doi: 10.1016/j.jecp.2012.08.002

Mayringer, H., and Wimmer, H. (2003). *Salzburger Lesescreening (SLS 1-4).* Bern: Hans Huber.

Mazzocco, M. M., and Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learn. Disabil. Res. Pract.* 20, 142–155. doi: 10.1111/j.1540-5826.2005.00129.x

Mazzocco, M. M., Feigenson, L., and Halberda, J. (2011a). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Dev.* 82, 1224–1237. doi: 10.1111/j.1467-8624.2011.01608.x

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

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., and Nuerk, H.-C. (2011). Early place-value understanding as a precursor for later arithmetic performance—A longitudinal study on numerical development. *Res. Dev. Disabil.* 32, 1837–1851. doi: 10.1016/j.ridd.2011.03.012

Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2009). Children's early mental number line: logarithmic or decomposed linear? *J. Exp. Child Psychol.* 103, 503–515. doi: 10.1016/j.jecp.2009.02.006

Mussolin, C., Mejias, S., and Noël, M.-P. (2010). Symbolic and non-symbolic number comparison in children with and without dyscalculia. *Cognition* 115, 10–25. doi:10.1016/j.cognition.2009.10.006

Nuerk, H.-C., Kaufmann, L., Zoppoth, S., and Willmes, K. (2004). On the development of the mental number line: More, less, or never holistic with increasing age? *Dev. Psychol.* 40, 1199–1211. doi: 10.1037/0012-1649.40.6.1199

Petermann, F., and Petermann, U. (2008). *Hamburg-Wechsler-Intelligenztest für Kinder IV (HAWIK-IV).* Bern: Huber.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012

Pixner, S., Moeller, K., Hermanova, V., Nuerk, H.-C., and Kaufmann, L. (2011). Whorf reloaded: language effects on nonverbal number processing in first grade—A trilingual study. *J. Exp. Child Psychol.* 108, 371–382. doi: 10.1016/j.jecp.2010.09.002

Pixner, S., Moeller, K., Zuber, J., and Nuerk, H.-C. (2009). Decomposed but parallel processing of two-digit numbers in 1st graders. *Open Psychol. J.* 2, 40–48.

Reeve, R., Reynolds, F., Humberstone, J., and Butterworth, B. (2012). Stability and change in markers of core numerical competencies.

J. Exp. Psychol. Gene 141, 649–666. doi: 10.1037/a0027520

Rousselle, L., and Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: a comparison of symbolic vs non-symbolic number magnitude processing. *Cognition* 102, 361–395. doi: 10.1016/j.cognition.2006.01.005

Rubinsten, O., and Henik, A. (2006). Double dissociation of functions in developmental dyslexia and dyscalculia. *J. Educ. Psychol.* 98, 854–867. doi: 10.1037/0022-0663.98.4.854

Rubinsten, O., Henik, A., Berger, A., and Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *J. Exp. Child Psychol.* 81, 74–92. doi: 10.1006/jecp.2001.2645

Schleifer, P., and Landerl, K. (2011). Subitizing and counting in typical and atypical development. *Dev. Sci.* 14, 280–291. doi: 10.1111/j.1467-7687.2010.00976.x

Schwarz, W., and Ischebeck, A. (2003). On the relative speed account of number-size interference in comparative judgments of numerals. *J. Exp. Psychol. Hum. Percep. Perform.* 29, 507–522. doi: 10.1037/0096-1523.29.3.507

Sekuler, R., and Mierkiewicz, D. (1977). Children's judgements of numerical equalitiy. *Child Dev.* 48, 630–633. doi: 10.2307/1128664

Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi: 10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–250. doi: 10.1111/1467-9280.02438

Siegler, R. S., and Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers' numerical understanding. *J. Educ. Psychol.* 101, 545–560. doi: 10.1037/a0014239

Trick, L. M., and Pylyshyn, Z. W. (1993). What enumeration studies can show us about spatial attention: evidence for limited capacity preattentive processing. *J. Exp. Psychol. Hum. Percep. Perform.* 19, 331–351. doi:10.1037/0096-1523.19.2.331

Vetter, P., Butterworth, B., and Bahrami, B. (2011). A candidate for the attentional bottleneck:

set-size specific modulation of the right TPJ during attentive enumeration. *J. Cogn. Neurosci.* 23, 728–736. doi: 10.1162/jocn.2010.21472

Von Aster, M., and Shalev, R. S. (2007). Number development and developmental dyscalculia. *Dev. Med. Child Neurol.* 49, 868–873. doi: 10.1111/j.1469-8749.2007.00868.x

Wilson, A. J., and Dehaene, S. (2007). "Number sense and developmental dyscalculia," in *Human Behavior, Learning, and the Developing Brain: Atypical Development,* eds D. Coch, G. Dawson, and K. Fischer (New York, NY: Guilford Press), 212–238.

World Health Organization. (2010). International Classification of Diseases (ICD-10). Retrieved from http://www.who.int/classifications/icd

Xu, F., and Arriaga, R. I. (2007). Number discrimination in 10-month-old infants. *Br. J. Dev. Psychol.* 25, 103–108. doi: 10.1348/026151005X90704

Xu, F., and Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition* 74, B1–B11. doi: 10.1016/S0010–0277(99)00066-9

Zimmermann, P., Gondan, M., and Fimm, B. (2002). *KITAP: Testbatterie zur Aufmerksamkeitsprüfung für Kinder.* Herzogenrath: Psytest.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Contributions from specific and general factors to unique deficits: two cases of mathematics learning difficulties

Vitor G. Haase[1,2]*, Annelise Júlio-Costa[1,2], Júlia B. Lopes-Silva[1,2], Isabella Starling-Alves[1], Andressa M. Antunes[1], Pedro Pinheiro-Chagas[3,4] and Guilherme Wood[5]

[1] Developmental Neuropsychology Laboratory, Department of Psychology, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[2] Programa de Pós-graduação em Saúde da Criança e do Adolescente, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[3] INSERM, U992, Cognitive Neuroimaging Unit, Gif sur Yvette, France
[4] CEA, DSV/I2BM, NeuroSpin Center, Gif sur Yvette, France
[5] Department of Neuropsychology, Institute of Psychology, Karl-Franzens-University of Graz, Graz, Austria

Mathematics learning difficulties are a highly comorbid and heterogeneous set of disorders linked to several dissociable mechanisms and endophenotypes. Two of these endophenotypes consist of primary deficits in number sense and verbal numerical representations. However, currently acknowledged endophenotypes are underspecified regarding the role of automatic vs. controlled information processing, and their description should be complemented. Two children with specific deficits in number sense and verbal numerical representations and normal or above-normal intelligence and preserved visuospatial cognition illustrate this point. Child H.V. exhibited deficits in number sense and fact retrieval. Child G.A. presented severe deficits in orally presented problems and transcoding tasks. A partial confirmation of the two endophenotypes that relate to the number sense and verbal processing was obtained, but a much more clear differentiation between the deficits presented by H.V. and G.A. can be reached by looking at differential impairments in modes of processing. H.V. is notably competent in the use of controlled processing but has problems with more automatic processes, such as nonsymbolic magnitude processing, speeded counting and fact retrieval. In contrast, G.A. can retrieve facts and process nonsymbolic magnitudes but exhibits severe impairment in recruiting executive functions and the concentration that is necessary to accomplish transcoding tasks and word problem solving. These results indicate that typical endophenotypes might be insufficient to describe accurately the deficits that are observed in children with mathematics learning abilities. However, by incorporating domain-specificity and modes of processing into the assessment of the endophenotypes, individual deficit profiles can be much more accurately described. This process calls for further specification of the endophenotypes in mathematics learning difficulties.

Keywords: endophenotype, mathematics learning difficulties, number sense, verbal numerical representations, phonological processing, dyslexia

INTRODUCTION

The cognitive underpinnings of arithmetic are highly complex (Rubinsten and Henik, 2009). One proposal is that arithmetic requires three types of symbolic and nonsymbolic number representations (Dehaene, 1992). The most basic form of numerical representation is nonsymbolic, analogic and approximate and corresponds to the number sense or the ability to discriminate numerosities. This ability can be described by Weber–Fechner's law, which measures the precision of the internal representation of numbers (Moyer and Landauer, 1967; Izard and Dehaene, 2008; Piazza, 2010). Precise numerical magnitude representations are related to phonologically and orthographically coded verbal numerals and visually based Arabic numerals (Dehaene and Cohen, 1995).

The number sense acuity is predictive of math achievement in both typical (Halberda et al., 2008; Mazzocco et al., 2011a) and disabled individuals (Piazza et al., 2010; Mazzocco et al., 2011b). Moreover, general cognitive resources are also involved in number processing, and calculations involve visuospatial abilities (Venneri et al., 2003), finger gnosias (Costa et al., 2011), phonological processing (De Smedt and Boets, 2010; De Smedt et al., 2010), working memory and executive functions (Camos, 2008; Pixner et al., 2011; Zheng et al., 2011).

The phenotypic presentation of mathematics learning disability and developmental dyscalculia (DD) is heterogeneous and includes a combination of the cognitive mechanisms that underlie arithmetic (Geary, 1993; Wilson and Dehaene, 2007). Because there are no consensual cognitive or biological markers, DD is operationally defined as persistent and severe difficulties in learning math in children of normal intelligence, that cannot be attributed to neurosensory impairment, sociodemographic, and emotional factors, or lack of adequate

educational experiences (American Psychiatric Association, 2000; World Health Organization, 2011). The nosological complexity of DD is compounded by its frequent comorbidity with other disorders, such as dyslexia (Landerl and Moll, 2010) and attention-deficit-hyperactivity disorder (ADHD, Gross-Tsur et al., 1996). Comorbidity can be explained by chance co-occurrences or by shared underlying mechanisms. The present evidence is still insufficient to decide about the role of comorbidity in characterizing DD (Rubinsten and Henik, 2009).

One possible way to solve the conundrum of DD's nosological validity is to consistently characterize implicated cognitive mechanisms as endophenotypes, in other words, as intermediate constructs between the interacting environmental and genetic etiologies and the phenotypic expression (Bishop and Rutter, 2009). A reliable endophenotype of number sense impairment has been gradually emerging (Piazza et al., 2010; Mazzocco et al., 2011b). However, restricting the definition of DD to individuals with more basic number processing impairments related to a number sense or number module (Reigosa-Crespo et al., 2012) would exclude from the domain of coverture of DD children and adolescents whose math learning difficulties could be persistent and of varying degrees of severity but associated with other cognitive mechanisms, such as phonological processing disorders (De Smedt and Boets, 2010).

Moreover, cognitive mechanisms that underlie math achievement and are potentially implicated in math learning difficulties could be classified as domain-specific or domain-general (Butterworth and Reigosa, 2007). Math-specific cognitive mechanisms include number sense (e.g., symbolic and nonsymbolic number comparison and estimation, number line estimation) and knowledge of the number system (Cowan and Powell, 2013). Domain-general mechanisms associated with math achievement and underachievement include phonological processing (Hecht et al., 2001), intelligence, processing speed, working memory, and executive functions (Cowan and Powell, 2013). It is increasingly recognized that DD can thus be characterized as primary, associated with number sense deficits, or secondary, associated with domain-general factors (Price and Ansari, 2013, for similar conceptions, see also Rubinsten and Henik, 2009; Reigosa-Crespo et al., 2012).

We argue that, in addition to being influenced by primary and secondary cognitive factors, the achievement profile of kids who struggle to learn math could also be affected by the nature of the information processing strategy that is deployed. An important research tradition in cognitive psychology, which dates back at least to Shiffrin and Schneider (1977), distinguishes between automatic (data-driven, bottom–up, effortless) and controlled (concept-driven, top–down, effortful) processing (Hasher and Zacks, 1979; Logan, 1988; Birnboim, 2003).

Evidence is still accumulating and is often inconsistent, but there are data that support impairments of both automatic and controlled processing in math learning difficulties. Impairments in the rapid automatized naming (RAN) of numbers (Bull and Johnston, 1997), a lack of the congruency effect in the number-size interference task (Rubinsten and Henik, 2005), and impairment in symbolic (with sparing of nonsymbolic) number comparisons (Rousselle and Noël, 2007) have been interpreted

as evidence for an automatization deficit in DD. Impairments of several subcomponents of the central executive in DD have often been described (Bull and Scerif, 2001; van der Sluis et al., 2004; Geary et al., 2007; Raghubar et al., 2010, see also Kaufmann et al., 2004; de Visscher and Noël, 2013). This literature indicates that math achievement could be associated with both domain-specific and domain-general cognitive factors. Moreover, these two dimensions could interact with different modes or strategies of information processing according to the nature of the task.

In general, it is possible to say that researchers agree as to the cognitive factors that are implicated in math learning difficulties. Disagreement arises when the relative importance of each factor or their possible interactions or lack of interaction are considered. One possibility is a multiple-deficit model, according to which math learning difficulties are the epigenetic outcome of multiple interacting mechanisms (Cowan and Powell, 2013). Another possibility is that different types of DD are explained by impairments in different non-interacting endophenotypes. One of the most important endophenotypes that is implicated in dyscalculia is a number sense or a number module deficit (Reigosa-Crespo et al., 2012). Single-case studies of individuals with math learning difficulties could constitute an opportunity to test these concurrent models of cognitive impairments in dyscalculia.

Although not without its critics (Thomas and Karmiloff-Smith, 2002), the logic of double-dissociation in cognitive neuropsychology has also been applied in the context of developmental disorders, to more specifically characterize the endophenotypes that are implicated (Temple, 1997; Temple and Clahsen, 2002; White et al., 2006a,b; de Jong et al., 2006, 2009). In cognitive neuropsychology, it is generally assumed that if two cognitive processes double-dissociate or present complementary patterns of spared and impaired functions in two different patients, then this pattern is an indication of different underlying neural substrates (Temple, 1997).

A possible double-dissociation in the field of learning disabilities is the case of the underlying cognitive mechanisms of DD and dyslexia. Evidence indicates that children with DD could be selectively impaired in number sense tasks, while dyslexia impairs phonological processing (Rubinsten and Henik, 2006; Landerl et al., 2009). Analysis has been performed on a series of single-case-generated evidence that is compatible with this interpretation (Tressoldi et al., 2007). The sole occurrence of DD and the sole occurrence of dyslexia, when associated with different cognitive profiles, suggest that these two disorders constitute distinct entities. At least in certain cases, the co-occurrence of DD and dyslexia could represent a true comorbidity, without a shared etiopathogenic variance (Landerl and Moll, 2010).

Double-dissociation logic has also been used to refine the phenotype of DD, characterizing subtypes that are related to impairments in specific cognitive components. A double dissociation has been observed in Arabic number processing. A case described by Temple (1989) presented a specific difficulty in reading Arabic numbers. The opposite difficulty of writing Arabic numbers was found by Sullivan et al. (1996). Similar to what is observed in adults with acquired acalculia, Temple (1991) demonstrated the existence of a double dissociation between procedural calculation impairment and a fact retrieval deficit. Specific fact

retrieval deficits were later corroborated by Temple and Sherwood (2007) in a group study. Two additional single-case studies described specific impairments in math facts retrieval, uncovering a role for executive function and automatization in the deficits (Kaufmann, 2002; Kaufmann et al., 2004; de Visscher and Noël, 2013). Moreover, more complex interactions between magnitude processing and procedural knowledge also can be observed in the carry over operation when solving addition problems (Klein et al., 2010). A number sense deficit impairing cardinality and sparing ordinality was observed in an earlier case described by Ta'ir et al. (1997).

This line of reasoning suggests, then, that single-case studies that use double-dissociation logic could play an important role in clearing the complexity that underlies phenotypic manifestations of DD and in establishing the relevant endophenotypes. Investigations on the number sense endophenotype using contemporary experimental measures are missing in the single-case literature. In this study, the aim is to contrast the patterns of cognitive deficits in two children at approximately 10 years of age with persistent math learning difficulties that are associated with distinct cognitive profiles. H.V., a 9-year-old girl, has math learning difficulties in the context of number sense inaccuracy, while G.A, a 10-year-old boy, presents math difficulties that are associated with developmental dyslexia and a phonological processing disorder. Neither of the children fulfilled the criteria for a more severe math learning disorder or disability. Instead, they were classified as having math learning difficulties, in other words, a performance below the 25th percentile on a standardized achievement test (Mazzocco, 2007). Performance on the Arithmetic subtest of the WISC-III was also not impaired in either of the children. Notwithstanding spared psychometric performance on achievement and intelligence tests, these two children presented persistent difficulties in specific domains of arithmetic, which were severe enough to cause low grades and to justify clinical referral.

The two cases were considered for analysis because of the comparable ages, similar sociodemographic backgrounds, normal or above average intelligence and impairment patterns that were suggestive of specific deficits in math learning difficulties. Standard neuropsychological assessment revealed specific impairments in the number sense in H.V. and in phonological processing in G.A. A more detailed assessment followed these observations.

Both domain-general and domain-specific cognitive mechanisms were included in the assessment (Butterworth and Reigosa, 2007; Cowan and Powell, 2013). Specific math assessment was based on two widely used cognitive models (McCloskey et al., 1985; Dehaene and Cohen, 1995). In the numerical domain, the following assessments were performed: numerical transcoding, calculation, simple word problems and the approximate number system (ANS).

Selection of domain-general assessments included the following functions: general intelligence (Deary et al., 2007), working memory (Geary et al., 2007; Raghubar et al., 2010), and executive functions (van der Sluis et al., 2004). Moreover, we used both non-numerical (Victoria Stroop, Strauss et al., 2006) and numerical stimuli (Five-digits Test, Sedó, 2007) when testing executive functions and interference (see the rationale in Raghubar et al.,

2010). Some aspects of our assessment protocol deserve further discussion. Phonological processing has been implicated in math learning (Hecht et al., 2001), mostly in the context of developmental dyslexia. A specific subtype of verbal dyscalculia has even been proposed (Wilson and Dehaene, 2007). Notwithstanding its theoretical plausibility, there is scarce evidence for a visuospatial subtype of dyscalculia (Geary, 1993; Wilson and Dehaene, 2007). Impairment of more executive aspects of visuospatial processing in math achievement has been reported, mostly in the context of the so-called nonverbal learning disability (Venneri et al., 2003). Wilson and Dehaene (2007) consider the possibility that impairments in the ANS and deficits in visuospatial attention could constitute two different subtypes of dyscalculia. It is important then to assess visuospatial and visuo-constructional abilities to check for the possibility of a nonverbal learning disability (Venneri et al., 2003; Fine et al., 2013). Finally, assessment of finger gnosias and motor dexterity were obtained because of their association with math learning difficulties (Costa et al., 2011; Lonnemann et al., 2011). Finger gnosias can underlie finger counting, which is an important offloading mechanism that liberates working memory resources at the beginning of formal math learning (Costa et al., 2011). Motor impairment could provide clues regarding the presence of minor brain insult (Denckla, 1997, 2003; Batstra et al., 2003).

METHODS

Considering the hypothesis that modes of information processing interact with the domain-specificity of stimuli in the genesis of learning difficulties, we employed tasks assessing automatic and controlled processing in both general and math-specific domains. General automatic processing was assessed using RAN of colors in the Victoria Stroop test. Numerical automatic processing was assessed by means of RAN of digits and speeded counting in the Five-digits Tests, nonsymbolic and symbolic number comparison tasks and by retrieval of arithmetic facts. Domain-general controlled processing was tapped by backward Corsi blocks span and the color-word interference phase of the Victoria Stroop test. Controlled processing in the numerical domain was evaluated with the backward Digit span and Inhibtion and Switching tasks of the Five-digits Test, as well as by word problems and working memory-dependent items in the numerical transcoding tasks. A simple reaction time task and the Nine-hole Peg Test were used to control, respectively for more basic aspects of alertness and motoric function.

CASE REPORTS

H.V. and G.A. were selected from cases at an outpatient facility for mathematical learning disabilities in Belo Horizonte, Brazil. Parents gave their written informed consent. In addition, informed consent was orally obtained from the children. Anamnestic information was obtained from the mothers of the two children.

H.V.

H.V. is a well-adjusted girl from a middle-class and supportive family, attending the third grade at a private school. She had just completed 9 years of age by the time of evaluation. H.V. had

difficulties in telling time on analogic and digital displays and estimating/comparing object sets (e.g., telling if a bookshelf had more or fewer books than another). She struggled to learn the math facts, to understand the place-value system and to solve word math problems. She uses fingers as a support to perform even the most simple additions and subtractions. Her learning difficulties are highly specific to math because her intelligence and achievement in other domains are above the average expected for her age. No major developmental problems were reported.

G.A.

G.A. is a well-adjusted boy from a middle-class and supportive family, who was 10 years and 2 months at the time of the neuropsychological assessment. He was attending the third grade at a public school. During his infancy, G.A. was submitted to several ear canal draining procedures that were related to recurrent otitis media. After the last surgery, his hearing and speech improved. His hearing is now normal and he was re-evaluated by a speech therapist who confirmed he has already improved from his previous difficulties.However, occasionally, he still mispronounces some of the more complex words, those that are less frequent and multi-syllable words that have consonantal clusters. G.A. was referred due to early and persistent difficulties with reading/spelling and math. His reading/spelling difficulties are severe. His math difficulties are milder but are also persistent and are mostly related to word problem solving. Clinically, G.A. presents difficulties with attention. A tentative diagnosis of ADHD was made by another clinician.

PROCEDURES

First, a general neuropsychological assessment was conducted, and the performances of both H.V. and G.A. were compared to available published norms. **Table 1** lists the neuropsychological tests and their sources. Afterward, an experimental study was conducted to specifically investigate math cognition in both cases. In the experimental investigation, the performances of H.V and G.A. were compared to two control groups that were individually matched by gender, educational level, age, and socioeconomic status. In Brazil, the type of school is an important indicator of socioeconomic status because private schools generally offer better instruction than public schools (Oliveira-Ferreira et al., 2012). For this reason and because of the age differences between the two patients, separate control groups were used for the comparisons. The controls were selected among the participants of a population-based research project on math learning difficulties that was approved by the local ethics review board. Parents gave written informed consent, and the children gave their oral consent.

The test performance of both cases was compared either to normed values, in the general neuropsychological assessment, or to the reference given by their individually selected control groups, in the math-cognitive assessment. Different statistical procedures that were based on psychometric single-case analysis (Huber, 1973; Willmes, 1985), one person vs. small sample comparisons (Crawford et al., 2010) and criterion-oriented methods (Willmes, 2003), were employed in these comparisons.

H.V.'s performance was compared to that of a group of 8 girls [mean age = 113 (SD = 3) months] from 3rd grade of a private

school in Belo Horizonte, Brazil. All of them had intelligence performance that was well above the mean (percentile ranks in the Raven's Colored Progressive Matrices ranged from 70 to 95) and no learning difficulties. G.A.'s performance, in turn, was compared to that of 17 boys [mean age = 117 (SD = 4) months] from the 3rd grade of two public schools in Belo Horizonte, Brazil. The percentile ranks in the Raven' Colored Progressive Matrices of this control group ranged from 50 to 99, which was comparable to that of G.A.'s.

INSTRUMENTS

In the following section, the more specific cognitive tests and tasks will be described in greater detail.

Brazilian school achievement test (TDE; Stein, 1994)

The TDE is a standardized test of school achievement (Oliveira-Ferreira et al., 2012) and comprises arithmetic, single-word spelling, and single-word reading. Specific norms are provided for school-age children between the second and seventh grade. Reliability coefficients (Cronbach α) of TDE subtests are 0.87 or higher. Children are instructed to work on the problems to the best of their capacity but without time limits.

Nine-hole peg test (9-HPT, Poole et al., 2005)

The 9-HPT is a timed test in which nine pegs should be inserted and removed from nine holes in the pegboard with the dominant and non-dominant hand. The pegboard is placed horizontally in front of the child, in such a way that the compartment that contains the pegs is on the side of the hand to be tested, while the compartment with the holes is on the contralateral side. Children must pick up one peg at a time. The test is performed two times with each hand, with two consecutive attempts with the dominant hand followed immediately by two consecutive attempts with the non-dominant hand. The scores were calculated based on the mean time for each hand.

Handedness ascertainment

Lateral preference was investigated by means of tasks that examine the ocular, hand, and foot preference based on Lefèvre and Diament (1982). The child was instructed to look through a hole, to kick and to throw a ball, three times each. The result was given by the side the child had chosen most of the time.

Right–left orientation test

This test is based on Dellatolas et al. (1998). It has 12 items of right and left body part recognition that involves simple commands regarding the child's own body, double commands (direct and crossed) toward the child's body, and pointing commands to single lateral body parts of an opposite-facing person. The score system is based on the number of correctly pointed parts of the body. Internal consistency was assessed with the Kuder–Richardson reliability coefficient, which was high (KR-20 = 0.80) (Costa et al., 2011).

Finger localization task

This 24-item task was also based on Dellatolas et al. (1998), and it was used to assess finger gnosia. It consists of three parts: (1) localization of single fingers touched by the examiner with the

Table 1 | Neuropsychological instruments.

Domain	Test	References
Psychosocial functioning	CBCL—Child Behavior Checklist responded to by parents	Achenbach et al., 2008; Rocha et al., 2012
School achievement	TDE—Teste de Desempenho Escolar (School Achievement Test)	Stein, 1994; Oliveira-Ferreira et al., 2012
Intelligence	Raven's colored progressive matrices	Angelini et al., 1999
	Wechsler intelligence scale for children 3° ed.	Figueiredo, 2002
Motor dexterity	9-HPT: Nine-hole peg test	Poole et al., 2005
Visuospatial abilities	Copy of the Rey–Osterrieth complex figure	Oliveira, 1999
Short-term and working memory	Corsi blocks	Santos et al., 2005
	WISC-III digits	Figueiredo and do Nascimento, 2007
	Auditory consonantal trigrams	Vaz et al., 2010
Phonological processing	Phoneme elision task	Lopes-Silva et al., 2014
	Pseudoword repetition	Santos and Bueno, 2003
	Pseudoword reading	Same stimuli as in Santos and Bueno (2003)
Executive functions	Color-word interference in the Victoria stroop	Charchat-Fichman and Oliveira, 2009
	Five-digits test	Sedó, 2004

hand visible (two trials on each hand); (2) localization of single fingers touched by the examiner with the hand hidden from view (four trials on each hand); and (3) localization of pairs of fingers simultaneously touched by the examiner with the hand hidden from view (six trials on each hand). A total score (that ranged from 0 to 12) was calculated for each child as well as the total score, which was the sum of the total from both hands. The internal consistency of this task is high (KR-20 = 0.79) (Costa et al., 2011).

Phoneme elision task

This test is a widely accepted measure of phonemic awareness (Wagner and Torgesen, 1987; Castles and Coltheart, 2004; Hulme et al., 2012; Melby-Lervåg et al., 2012). The child listens to a word and is expected to say how it would be if a specified phoneme were deleted. (e.g., "*filha*" without /f/ is "*ilha*" in English it would be "cup" without /k/ is "up"). The test comprises 28 items: in 8 of them, the child must delete a vowel, and in the other 20, a consonant. The consonants to be suppressed varied according to the place and manner of articulation. The phoneme to be suppressed could be in different positions of the words, which ranged from 2 to 3 syllables. The internal consistency of the task is 0.92 (KR-20 formula) (Lopes-Silva et al., 2014).

Victoria stroop task (*Charchat-Fichman and Oliveira, 2009*)

The Victoria Stroop task is a measure of executive function (Strauss et al., 2006). The subject is presented with three cards, each containing six rows of four items. In the first card (color), the task is to name quickly the color of 24 rectangles, which can be green, yellow, blue, or red. In the second card (word), the task is to name the colors of common words printed in green, yellow, blue, or red, ignoring their verbal content. On the third card, the stimuli are color names that are printed in an incongruent color that is never the same color as the word that is printed. The task is to name the color in which the word is printed (e.g., when the word "blue" is printed in red, the subject must say "red"). For each of the three conditions, the time to complete the naming of all of the stimuli was recorded. Additionally, the interference score

(Stroop-Effect) was calculated as the quotient between the time score for the incongruent (third card) and the color (first card) conditions.

Five digits test

The Five Digits Test was validated and standardized in Spanish and English by Sedó (2004, 2007) as a measure of speeded counting, Arabic number reading, and inhibition and set shifting. Similar sets of stimuli are used across tasks. Automatic processing is assessed through speeded tasks of counting randomly presented star sets (up to five) and reading Arabic digits (up to five). Controlled processing is assessed through inhibition and set-shifting tasks. In the inhibition task (choosing), the child must count the number of Arabic digits instead of reading them. In the set-shifting condition (switching), the child switches from counting the number of Arabic digits in most trials to reading them when a frame surrounds the stimulus set.

The numeric and arithmetic tasks for the experimental study have been employed in previous investigations (Costa et al., 2011; Ferreira et al., 2012; Júlio-Costa et al., 2013) and are described below.

Simple reaction time

The computerized simple RT task is a visual detection task that is used to control possible differences in the basic processing speed that is not related to numerical tasks. In this task, a picture of a wolf (height 9.31 cm; length = 11.59 cm) is displayed in the center of a black screen for a maximum time of 3000 ms. The participant is instructed to press the spacebar on the keyboard as fast as possible when the wolf appears. Each trial was terminated with the first key press. The task has 30 experimental trials, with an intertrial interval that varies between 2000, 3500, 5000, 6500, and 8000 ms.

Nonsymbolic magnitude comparison task

In the nonsymbolic magnitude comparison task, the participant was instructed to compare two simultaneously presented sets of dots and to indicate which set contained the larger number (see **Figure 1**). Black dots were presented on a white circle over a black

FIGURE 1 | Nonsymbolic magnitude comparison task.
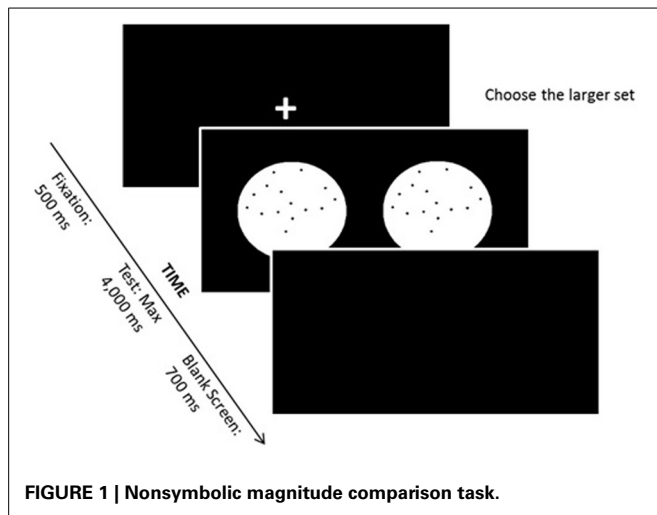
background. On each trial, one of the two white circles contained 32 dots (reference numerosity), and the other circle contained 20, 23, 26, 29, 35, 38, 41, or 44 dots. Each magnitude of dot sets was presented 8 times. The task comprised 8 learning trials and 64 experimental trials. Perceptual variables were randomly varied such that in half of the trials, the individual dot size was held constant, while in the other half, the size of the area occupied by the dots was held constant (see exact procedure descriptions in Dehaene et al., 2005). The maximum stimulus presentation time was 4000 ms, and the intertrial interval was 700 ms. Between each trial, a fixation point appeared on the screen—a cross, printed in white, with 30 mm in each line. If the child judged that the right circle presented more dots, then a predefined key localized on the right side of the keyboard should be pressed with the right hand. In contrast, if the child judged that the left circle contained more dots, than a predefined key on the left side had to be pressed with the left hand.

Symbolic magnitude comparison task

In the symbolic magnitude comparison task, Arabic digits from 1 to 9 were presented on the computer screen (height = 2.12 cm; length = 2.12 cm). The visual angle of the stimuli was 2.43° in both the vertical and horizontal dimensions. Children were instructed to compare the stimuli with the reference number 5. Digits were presented in white on a black background. If the presented number was smaller than 5, the child had to press a predefined key on the left side of the keyboard, with the left hand. If the stimulus was higher than 5, then the key to be pressed was located at the right side and should be pressed with the right hand. The number 5 was never presented on the computer screen. Numerical distances between stimuli and the reference number (5) varied from 1 to 4, each numerical distance being presented the same number of times. Between trials, a fixation point of the same size and color of the stimuli was presented on the screen. The task comprised 80 experimental trials. The maximum stimulus presentation time was 4000 ms, and the intertrial interval was 700 ms.

Simple calculation

This task consisted of addition (27 items), subtraction (27 items), and multiplication (28 items) operations for individual applications, which were printed on separate sheets of paper. Children were instructed to answer as fast and as accurately as they could, with the time limit per block being 1 min. Arithmetic operations were organized at two levels of complexity and were presented to children in separated blocks: one consisted of simple arithmetic table facts and the other consisted of more complex facts. Simple additions were defined as those operations that had results of below 10 (i.e., 3 + 5), while complex additions had results between 11 and 17 (i.e., 9 + 5). Tie problems (i.e., 4 + 4) were not used for addition. Simple subtraction comprised problems in which the operands were below 10 (i.e., 9 − 6), while for complex subtractions, the first operand ranged from 11 to 17 (i.e., 16 − 9). No negative results were included in the subtraction problems. Simple multiplication consisted of operations that had results of below 25 and that had the number 5 as one of the operands (i.e., 2 × 7, 5 × 6), while for the complex multiplication, the result of the operands ranged from 24 to 72 (6 × 8). Tie problems were not used for multiplication. Reliability coefficients were high (Cronbach's $\alpha > 0.90$).

Simple word problems

Twelve arithmetical word problems were presented to the child on a sheet of paper while the examiner read them aloud simultaneously to avoid a reading proficiency bias. There were six addition and six subtraction items, all of them with single-digit operands and results that ranged from 2 to 9 (i.e., "Annelise has 9 cents. She gives 3 to Pedro. How many cents does Annelise have now?"). The child had to solve the problems mentally and write the answer down in Arabic format as quickly as possible, and the examiner registered the time that was taken for each item. Cronbach's α of this task was 0.83.

Arabic number reading task

Twenty-eight Arabic numbers printed in a booklet were presented one at a time, to the children, who were instructed to read them aloud. The item set consists of numbers up to 4 digits (3 one-digit numbers, 9 two-digit numbers, 8 three-digit numbers, and 8 four-digit numbers). There were 12 numbers that could be lexically retrieved, 5 numbers that required three transcoding rules according to the ADAPT model (Barrouillet et al., 2004) to be correctly read, 6 numbers with four rules and 5 numbers with more than five rules. The internal consistency of the task is 0.90 (KR-20 formula) (Moura et al., 2013).

Arabic number writing task

Children were instructed to write the Arabic form of dictated numbers. This task is composed of 40 items, with up to 4 digits (3 one-digit numbers, 9 two-digit numbers, 10 three-digit numbers, and 18 four-digit numbers). The one- and two-digit numbers were classified as "lexical items" (12 items), and the other 28 items were subdivided according to the number of transcoding rules based on the ADAPT model (Barrouillet et al., 2004; Camos, 2008). There were six numbers that require 3 rules, nine numbers

that require 4 rules, six numbers with 5 rules, five numbers with 6 rules, and two numbers with 7 rules. The internal consistency of this task is 0.96 (KR-20 formula) (Moura et al., 2013).

RESULTS

GENERAL COGNITIVE ASSESSMENT

Results of the CBCL reported by their respective mothers were in the normal range in all of the subscales (T-scores in the single subscales ranged from 37 to 45 in H.V. and from 36 to 54 in G.A. Scores above 70 are considered to be clinical). This finding indicates that both children have adequate levels of psychosocial functioning, according to their mothers. The results of the intelligence test are exhibited in **Figure 2**, while **Figure 3** depicts comparative results in the two cases for the general neuropsychological assessment compared to norms from the original publications. H.V. shows a performance in the upper bound of normal intelligence (Raven's $PR = 99$, $FSIQ = 120$, $VIQ = 116$, and $PIQ = 121$), and G.A. shows average intelligence (Raven's $PR = 75$, $FSIQ = 87$, $VIQ = 89$, and $PIQ = 89$).

Statistical comparisons between both children in the subtests that measure the verbal and performance IQs (Huber, 1973; Willmes, 1985) reveal significantly higher scores for H.V. in the subtest Information ($Z = 2.95$; $p = 0.016$), Similarities ($Z = 3.33$; $p = 0.004$), Arithmetic ($Z = 2.58$; $p = 0.05$), Vocabulary ($Z = 4.87$; $p = 0.00001$), Figure Assembly ($Z = 2.36$; $p = 0.01$) and Coding ($Z = 5.59$, $p = 0.000001$). These results disclose a general pattern of higher scores in H.V. than in G.A. regarding

tasks that demand more from verbal IQ but not as much regarding performance IQ.

Performance on the TDE (Brazilian School Achievement Test) was below the 25th percentile in both cases for Arithmetic. H.V.'s accuracy percentage was 29% (raw score = 11, grade mean = 16 grade, $SD = 3.39$) and G.A.'s was 36% (raw score = 14, grade mean = 16, grade $SD = 3.39$). The 25th percentile criterion is used as a lenient cut-off and is sensitive to math learning difficulties (Mazzocco, 2007; Landerl and Kölle, 2009; Landerl et al., 2009). Performance on the single word Reading and Spelling
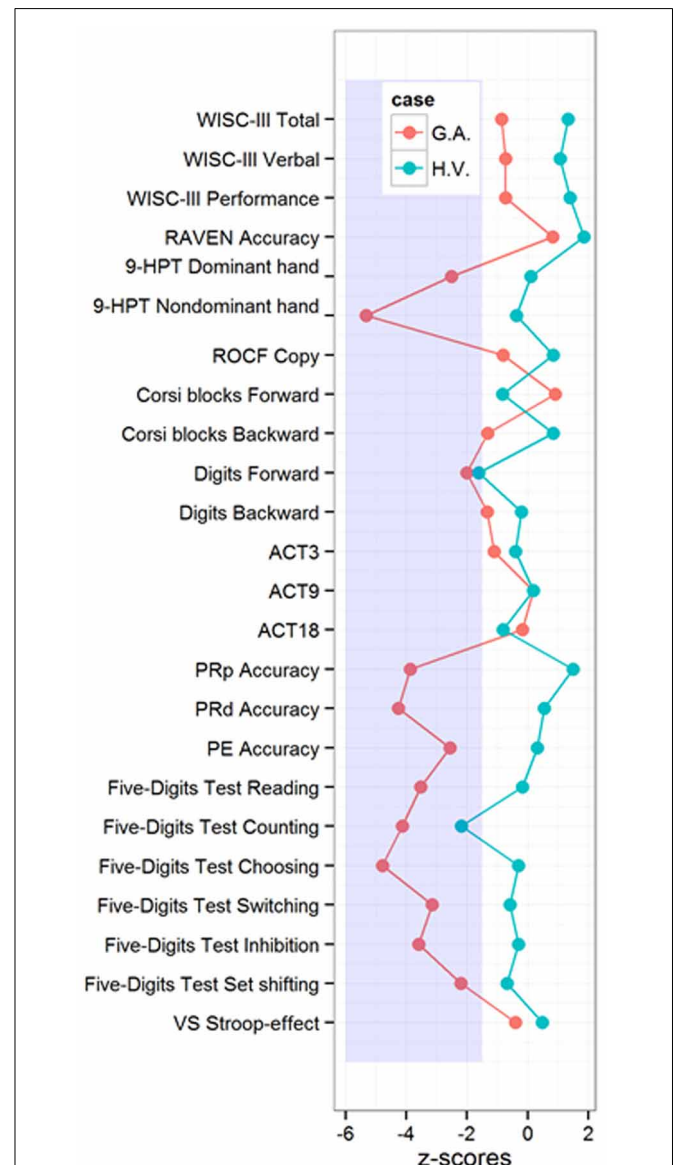
FIGURE 3 | H.V. and G. A. performances in intelligence, motor dexterity, visuospatial, short-term and working memory, executive functions, and phonological processing tasks. 9-HPT, Nine Hole Peg Test; ROCF, Rey-Osterrieth Complex Figure; ACT, Auditory consonant trigrams; PRp, Pseudoword Repetition; PRd, Pseudoword Reading; PE, Phoneme Elision; VS, Victoria Stroop. Clinical score < −1.5 *SD*.

FIGURE 2 | H.V. and G. A. performances in WISC-III. *Marked statistical significance at the level *p* < 0.001. Note: as H.V.'s standardized Block Design score was below the mean in the first assessment, this subtest was repeated two years later (gray dot). The new standardized score in Block Design was equal to 10.

subtests of the TDE were normal for H.V. and below the 25th percentile for G.A. G.A. solved 14 out of the 35 items of the Spelling subtest correctly. In some items, he excluded phonemes (especially /r/, regardless of its mode or place of articulation), and in others, he confused phonemes that have similar sounds (such as v/f; m/n; b/d; and s/c). He clearly presented a phonological writing pattern, but he still lacks the mastery of the alphabetical principle. In the Reading subtest, G.A. could read 55% of the single words (raw score = 39, grade mean = 64.75, SD = 4.67), and his reading was extremely slow. He struggled at reading consonant clusters.

Regarding motor dexterity in the 9-HPT, H.V. did not present any major difficulties, whereas G.A.'s score was on the adopted clinical range, which means that he was significantly slower than would be expected for his age range according to Poole et al. (2005) (**Figure 2**). Both children presented right hand dominance (Lefèvre and Diament, 1982) as well as normal right-left orientation (Dellatolas et al., 1998) and finger gnosias (Dellatolas et al., 1998) (**Figure 2**). Neither of the children presented visuospatial constructional deficits.

On the phonological processing tasks, G.A. was significantly worse on all of the tests that were used, while H.V. presented typical scores. G.A. presented difficulties in storing and reproducing pseudowords as well as in reading them. In addition, he was not able to grasp the grapheme-phoneme correspondence principle that is needed to perform the phoneme elision task.

Both children presented difficulties in the phonological short-term memory task (forward digit span), but in both cases, scores in the phonological working memory tests (backward order of the Digit Span as well as the Auditory Consonantal Trigrams) fell into the expected ±1.5 SD range (**Figure 2**). This specific difficulty on the forward order of the Digit Span was mild, and it can be attributed to attentional lapses (Strauss et al., 2006). G.A. presented a better performance on the forward order of the Corsi Blocks compared to the backward, and H.V. showed the opposite pattern. However, both of their spans were in accordance to what would be expected for their age range. The performance of both children was in the typical range for the Victoria Stroop task. G.A.'s performance was in the clinical range for all of the subtests of the Five-digits test, those that involve more automatic processing (speeded digit reading and counting) as well as those that require executive functioning (inhibition and shifting). H.V. presented only a specific impairment that involved counting skills on the Five-digits Test, which will be discussed in more detail below[1].

---

[1]After the neuropsychological assessment, both children initiated interventions based on cognitive-behavioral techniques to reduce math-anxiety symptoms and also to improve self-efficacy. Strategies such as problem-solving, self-monitoring, and self-reinforcement are coupled with errorless learning, allowing the kids to have experiences of academic success. Simultaneously, we also use instructional and training interventions that focused on number processing and arithmetic components that were considered to be impaired in each child. H.V. and G.A. have been participating in individual intervention programs for 4 semesters, 2 h a week. Their families also received counseling by means of a behavioral training program for one semester, once a week. During this time, H.V. has not obtained improvements in her number sense acuity, but she considerably improved in solving addition and multiplication

MATH COGNITIVE ASSESSMENT AND COMPUTER TASKS

Results of the computerized and math-cognitive tasks are shown in **Tables 2**, **3** for H.V. and in **Tables 4**, **5** for G.A. and their respective control groups.

In the simple reaction time task, H.V. did not show any impairment. In contrast, she responded faster than the average of her group. In the symbolic number task, the picture is different. Although H.V. was significantly slower than her control group, her response accuracy was slightly higher than that of controls in a type of speed-accuracy trade-off. Moreover, the performance of H.V. in the nonsymbolic task was markedly impaired in comparison with her control group. While the reaction times were comparable to the group average, the accuracy was very poor, especially for the more difficult numerical ratios. These deficits added to the picture that was formed by a speeded counting impairment in the Five-digits Test. The results of the number processing tasks suggest that there was a specific impairment in the number sense acuity in the presence of relatively spared numerical symbolic abilities.

H.V.'s performance was substantially impaired in complex addition and multiplication operations. Her performance was comparable to the control group in simple word problems (**Table 3**). H.V. can solve simple addition and subtraction operations as accurately as expected according to her age. In complex addition operations, H.V. presents more difficulties when compared to her control group. Interestingly, these difficulties could not be observed in complex subtraction tasks. Moreover, in comparison to controls, H.V. shows systematic difficulties when solving simple and complex multiplication operations, which can be interpreted as a more general deficit in fact retrieval. No deficits were observed in simple word problems with one-digit operands, the solution of which depends on text comprehension; these problems can be solved by counting procedures. She solved all of the problems correctly but took considerably more time to reach the correct results. Performance on number transcoding of three- and four-digit numerals was comparable to the control group (**Table 3**). These results are summarized in **Table 3**.

In the simple reaction time task, G.A. did not show any impairment but instead showed average performance (**Table 4**). In the symbolic number task, G.A. responded tendentially slower and much less accurately than his control group. In contrast, G.A. presented both average response latency and average accuracy in the nonsymbolic number comparison task. In the number processing tasks, G.A. experienced considerable difficulties in tasks that use the symbolic notation and verbal procedures, such as speeded counting, speeded digit reading, transcoding and symbolic magnitude comparison (up to nine). G.A.'s pattern of impairment in the math tasks contrasts with that of H.V. Difficulties in the symbolic number processing tasks in G.A. are at odds with a normal Weber fraction.

G.A.'s difficulties with the symbolic processing were also corroborated by his lower performance in the transcoding tasks.

---

problems. She has not automatized fact retrieval yet. H.V. does her homework with a pocket calculator. Initially, G.A. received training in text processing abilities and improved in arithmetical word problem solving. He also obtained substantial improvement in his transcoding abilities.

Table 2 | Descriptive data and comparison between the control groups and H.V., in the alertness and number sense tasks.

Domain assessed	Task	Controls (n = 08)		H.V.	Modified t-test	p	Z-CC	Estimated % pop. below H.V.
		Mean	SD					
Alertness	Simple manual reaction time	413.3	32.9	381.3	−0.92	0.39	−0.97	19
Number sense	**SYMBOLIC MAGNITUDE COMPARISON**							
	Response time*	771.6	154.8	1153.3	2.32	0.05	2.47	97
	Weber Fraction	0.27	0.17	0.12	−0.83	0.21	−0.88	22
	NON-SYMBOLIC MAGNITUDE COMPARISON							
	Response time*	1035.5	199.5	1003.2	−0.15	0.88	−0.16	44
	Weber fraction	0.29	0.06	0.42	2.24	0.04	2.17	4

*ZCC: magnitude effect index calculated by the difference between the scores of the control group and the single case with a 95% CI (Crawford et al., 2010); *time in milliseconds.*

Table 3 | Descriptive data and comparisons between control groups and H.V. in the Simple calculation, Simple word problems, and Verbal-Arabic transcoding tasks (df = 1).

Domain assessed	Task (total of items)	Controls (n = 08)		H.V.	X²	p
		Mean	SD			
Simple calculation	**BASIC ARITHMETIC OPERATIONS**					
	Simple addition (12)	11.88	0.35	9	1.30	0.25
	Complex addition (15)	13.88	2.23	7	5.45	0.02
	Simple subtraction (12)	10.50	1.69	11	<0.01	1.00
	Complex subtraction (15)	8.50	3.25	4	1.68	0.19
	Simple multiplication (15)	13.13	2.64	4	11.18	0.00
	Complex multiplication (13)	6.25	3.45	0	5.81	0.02
Simple word problems	Math word problems (12)	10.50	2.35	12	0.18	0.67
Verbal-Arabic transcoding	Arabic number writing task (40)	38.00	3.22	40	0.51	0.47
	Arabic number reading task (28)	27.75	0.71	28	<0.01	1.00

Table 4 | Descriptive data and comparison between control groups and G.A. in the alertness and number sense tasks.

Domain assessed	Task (total of items)	Controls (n = 17)		G.A.	Modified t-test	p	Z-CC	Estimated % pop. below G.A.
		Mean	SD					
Alertness	Simple manual reaction time*	423.8	82.3	447.9	0.29	0.39	0.29	39
Number sense	**SYMBOLIC MAGNITUDE COMPARISON**							
	Response time*	983.1	249.4	1344.9	1.41	0.09	1.45	9
	Weber fraction	0.21	0.14	0.78	3.96	<0.001	4.07	>99
	NON-SYMBOLIC MAGNITUDE COMPARISON							
	Response time*	1276.3	294.9	1038.2	−0.79	0.22	−0.81	39
	Weber fraction	0.28	0.10	0.21	−0.68	0.25	−0.70	25

*ZCC: magnitude effect index calculated by the difference between the scores of control group and single-case with a 95% CI (Crawford et al., 2010); *time in milliseconds.*

Table 5 | Descriptive data and comparison between control groups and G.A. in the Simple calculation, Simple word problems, and Verbal-Arabic transcoding tasks ($df = 1$).

Domain assessed	Task (total of items)	Controls ($n = 17$)		G.A.	X^2	p
		Mean	SD			
Simple calculation	**BASIC ARITHMETIC OPERATIONS**					
	Simple addition (12)	11.71	0.69	6	4.78	0.03
	Complex addition (15)	10.12	3.06	5	2.26	0.13
	Simple subtraction (12)	10.41	2.15	8	0.46	0.50
	Complex subtraction (15)	5.18	3.21	2	0.87	0.35
	Simple multiplication (15)	9.71	4.58	5	1.84	0.18
	Complex multiplication (13)	2.35	2.12	0	0.85	0.36
Simple word problems	Math word problems (12)	10.82	1.19	4	5.98	0.01
Verbal-Arabic transcoding	Arabic number writing task (40)	38.65	2.78	26	10.94	<0.001
	Arabic number reading task (28)	27.59	0.87	16	11.61	<0.001

In the number writing task, G.A. committed 14/40 errors. G.A. presented three lexical errors (all of them were related to phonological resemblance between the trial and the number written by him) and eleven syntactic ones (seven being related to adding internal zeros and four to deleting a digit). Fifty-two percent of his control group did not commit any error. From the eight children who did, one committed eleven errors, one presented five errors, one committed two errors and the other five children made only one single mistake.

The lexical mistakes by G.A. clearly have a phonological bias. In Portuguese, the numbers "three" and "six" sound very similar ("três" and "seis," respectively), as well as "seven hundred" and "six hundred" ("setecentos" and "seiscentos"). Moreover, the syntactic errors of G.A. always involved the addition principle (overwriting rule, Power and Dal Martello, 1990; Moura et al., 2013). G.A. wrote the number 643 as 646 and 4701 as 400601. His performance on the number reading test also corroborates his difficulties with place value understanding. He read the number "2000" as "two hundred" and "1013" as "one hundred thirteen." On two items, he decomposed the numbers: 567 was read as "five and sixty seven" and 5962 as "fifty nine and sixty two." Nevertheless, the mistakes made by G.A. cannot be easily attributed to a lack of knowledge of the rules of additivity in number transcoding. G.A. was able to transcode correctly five out of eleven complex numbers with syntactical zeros (e.g., "109," "902," "1060," "1002," and "7013") but failed to transcode numbers of comparable complexity ("101" ≥ 11, "1015" ≥ 10015, "2609" ≥ 20069, "4701" ≥ 40601, "1107" ≥ 2067, and "7105" ≥ 715). Therefore, the poor transcoding performance of G.A. is compatible with deficits in phonological representations combined with problems with concentration and monitoring capacity. Evidence for a deficit in knowledge about the structure of the Portuguese verbal number system was not obtained.

Difficulties with simple word problems were more severe. G.A. did not show any impairment in solving addition, subtraction and multiplication problems when compared to controls, except for a single result that indicated lower performance while solving simple addition tasks (**Table 5**). This pattern is consistent with the mother's report that G.A. acquired the arithmetic facts after struggling with them for a while. However, the verbal nature of G.A.'s difficulties becomes explicit again, when considering his attainment of simple word problems. From 12 problems, G.A. solved only 4 correctly, responding sometimes with absurd values, which suggested that he was guessing. His performance on word problems was almost six standard scores below that of the controls. In summary, the results of the math cognitive investigation suggest that G.A.'s difficulties in learning math can be attributable to his comorbid reading learning disability.

DISCUSSION

In the present study, we selected two cases that had relatively specific impairment patterns from an outpatient clinic for mathematical learning disorders and conducted a detailed neuropsychological and cognitive assessment with the aim of characterizing possible endophenotypes. Specificity of the impairments is corroborated by the fact that both children were of average or above average intelligence and did not present impairments in visuospatial and visuoconstructional processing, as assessed by the Rey figure copy and Block Design subtest of the WISC. In the following, we will discuss the extent to which the neuropsychological profile of H.V. and G.A. fitted specific endophenotypes, as predicted in the literature.

H.V.

Difficulties in H.V. are specific, severe and persistent and were restricted to an inaccurate number sense and to the acquisition of arithmetic facts, which reflected mostly on multiplication operations. H.V. is curious and motivated to learn, except for mathematics. H.V. has difficulties in memorizing even the simplest arithmetic facts, but she is highly skilled in finger counting. The single abnormally lower score observed in the general neuropsychological assessment was in the forward version of the digit span. An excellent performance was observed in reading-related phonological processing tasks, such as pseudoword repetition,

pseudoword reading and phonemic ellison. No abnormalities were observed in executive function tasks.

One might wonder why the performance of H.V. in the subtest "counting" of the Five-digits Test of executive functions was so low and discrepant from her general level of performance on this test. The subtest counting is a speeded task in which one has to count how many stars are printed on a series of cards that display sets of one up to five stimuli. The difficulties with the speeded counting of stars presented by H.V. reflect much more a deficit in the apprehension of nonsymbolic magnitude information under time constraints. This pattern contrasts with her resourceful use of strategies to compensate for her difficulties in other tasks that do not require nonsymbolic number processing. One of her favorite compensatory strategies for solving even the simplest arithmetic problems is finger counting. Once sufficient time is allowed, H.V. can find the correct response by finger counting. Her difficulties are accentuated in speeded tasks that require automatic retrieval.

The deficits in fact retrieval that are presented by H.V. cannot be attributed to a reduced capacity of verbal working memory or phonological awareness because H.V. shows high levels of competence in these two cognitive functions. However, the deficits in the numerical and arithmetic abilities of H.V. are compatible with generally imprecise or poor numerical representations: on the one hand, the deficits of H.V. in multiplication tasks suggest impairment in the retrieval of appropriate information from memory. On the other hand, the high value of the Weber fraction observed in nonsymbolic magnitude comparison suggests a very inaccurate ANS.

In contrast to our expectations, the profile of H.V. does not fit a typical endophenotype that is characterized by a number sense deficit (Wilson and Dehaene, 2007; Noël and Rousselle, 2011). Although H.V. presents low acuity in nonsymbolic magnitude comparison, this deficit is not present in the symbolic version of the task. More importantly, a substantial deficit in arithmetic operations—particularly in subtraction—was not observed. In contrast, H.V. presented some deficit in complex addition operations, but no sign of a deficit was observed in simple or complex subtraction operations. Moreover, a substantial deficit in multiplication operations (simple as well as complex) cannot be accounted for by a deficit in the number sense alone, but suggests the presence of difficulties for automatizing the retrieval of multiplication facts.

G.A.

G.A. presented persistent but milder difficulties in learning math in the context of developmental dyslexia with severe associated phonemic processing deficits. In the case of G.A., math learning impairments were observed in transcoding operations as well as in very simple one-digit word problems. G.A. presented deficits in all phonological processing tasks: digit span, pseudoword repetition and reading as well as in phoneme elision. Although his intelligence is normal, difficulties were also observed in motor dexterity and in all subtests of the five-digits procedure, both those tapping automatic (speeded counting, speeded digit reading) and those assessing controlled processing (inhibition, set shifting). Moreover, a borderline performance was also

observed in the forward and backward Digit and backward Corsi spans.

G.A. showed a less pronounced deficit in numerical and arithmetical abilities than H.V. The acuity of his representation of magnitude was comparable to controls, as measured by the nonsymbolic magnitude comparison task. In contrast, in the symbolic magnitude comparison task, G.A. committed many more errors and was marginally slower than his control group. Although G.A. presented lower levels of performance than controls in the simple addition operations, no other difference was observed in simple or complex addition, subtraction or multiplication operations. This pattern indicates that G.A. can retrieve from memory the correct responses to simple operations and employ the correct procedures to execute more complex addition and subtraction operations. However, in comparison to the controls, G.A. was much less successful when solving word problems. G.A. also presented substantially more difficulties in transcoding tasks in comparison to his peers, especially regarding phonological representations, concentration and monitoring capacity.

The profile of G.A. fits only partially a typical endophenotype that is characterized by a verbal and symbolic deficit. Although G.A. presents low acuity in symbolic magnitude comparison, simple word problems and impaired performance in transcoding tasks, this deficit does not extend to the retrieval of multiplication facts. It is still a matter of debate to what extent multiplication facts are stored in a typical verbal format (Varley et al., 2005; Benn et al., 2012). However, deficits in verbal numerical information processing have, very often, been associated with deficits in fact retrieval (De Smedt and Boets, 2010; De Smedt et al., 2010).

G.A. also presents severe problems with motor dexterity, which are assessed with the 9-HPT, which deserve consideration. Sensorimotor impairments are a frequent concomitant of specific learning disorders observed both in dyslexia (White et al., 2006a,b) and in dyscalculia (Costa et al., 2011; Lonnemann et al., 2011). Minor sensorimotor dysfunction was observed in 87% of dyslexic children with an IQ higher than 85 (Punt et al., 2010). In this context, they are not interpreted as a causal mechanism that is implicated in learning difficulties, but as markers or co-localizers of brain insult (Denckla, 1997, 2003; Batstra et al., 2003). Whatever the cause of G.A.'s present learning difficulties, it also impaired his neurological functions in a more widespread manner, as shown by the relatively severe reduction in motor dexterity. Because the motor difficulties were comparable in both hands, no inferences can be made regarding lateralization of the underlying pathological process, other than the left-hemisphere dysfunction that is connected to developmental dyslexia.

In our view, the sensorimotor deficits could be responsible for his deficits in other tasks as well. G.A.s performance in both the Block Design subtest and Rey's Figure copy were situated from 0.7 to 1 standard deviations below the mean, which suits his WISC-FSIQ of 90. Moreover, a qualitative assessment of G.A.'s performance in the Block Design subtest and Rey's Figure copy indicate that his relative difficulties originate from the motor dexterity and executive components that are mobilized to solve these tasks and do not reflect impairments in apprehension or reproduction of visuospatial configurations. Further corroboration of

these findings comes from the Raven. There, G.A. reached a score that was higher than average. In our view, such a level of performance on the Raven cannot be reached when simple visuospatial processing is impaired.

The difference between G.A.'s scores on the WISC and Raven can be attributed to an interaction between test and individual characteristics. Compared to the Raven, the WISC-III imposes greater demands on verbal and scholastic abilities. Performance on several WISC tasks is also time constrained. We believe that G.A.'s relatively lower performance on the WISC can be explained by his reading and academic difficulties as well as by impairment in motor dexterity and processing speed. This pattern is especially salient on the Coding subtest, which is the test that presents the worst performance. Difficulties with the Coding subtest can also be related to G.A.'s impairment with respect to the symbolic transcoding tasks (Strauss et al., 2006).

SPECIFIC DEFICITS IN AUTOMATIC vs. CONTROLLED NUMERICAL PROCESSING?

Comparisons of the endophenotypes as predicted by the current literature (Wilson and Dehaene, 2007; Noël and Rousselle, 2011) and the individual cases of H.V. and G.A. yield apparently frustrating results because the performance of H.V. and G.A. on the arithmetic tests partly contradicts the general expectation of more or fewer specific deficits in the number sense and verbal numerical representations, respectively. One possible interpretation of these results is that paradigmatic cases that regard specific endophenotypes can be very difficult to find. Although the initial assessment of H.V. and G.A. suggested number sense and verbal deficits, a more detailed examination revealed, in both cases, a less precise picture. Similar difficulties encountered by other authors (e. g., Tressoldi et al., 2007), suggest that only a small proportion of all of the cases of mathematics difficulties can reveal more pure forms of endophenotypes. This finding raises the question about the proportion of cases of mathematics difficulties that can actually be assigned with confidence to one or another subtype of this disorder. If it is low, then the general approach of endophenotypes might prove to be ineffective. Although our case design does not allow a direct investigation of this question, in this section, we will discuss one possible reason why endophenotypes can be indeed valuable in the investigation of mathematics difficulties.

One could propose that the severe deficits of H.V. solving multiplication problems while simultaneously being capable of solving complex subtraction problems are a result of compensatory strategies, such as finger counting. Finger counting could be more effective for subtraction than for multiplication operations because the multiplication operations usually have much higher numbers as the answers, which are much more difficult to reach by counting. Assuming that this reason explains H.V.'s performance, the discrepancy between her performance and the typical results that are expected according to the number sense endophenotype should be due to relatively trivial differences between prototypical profiles and individual cases, without more profound consequences for the refinement of the theoretical framework of mathematical learning disorders.

The same conclusion can be reached when analysing the discrepancy between G.A.'s performance and a verbal numerical endophenotype. Deficits in calculations should be expected, especially when the problems are more complex, rely more strongly on a verbal code, and the ability to use verbal number representations is as limited as in the case of G.A. However, this expectation was not confirmed by the results. Once more, one can attribute the discrepancy between the observed performance and typical endophenotypes to some individual compensatory resource, which is always plausible in individual cases and is frequently reported in clinical observations (Temple and Clahsen, 2002; Thomas and Karmiloff-Smith, 2002).

Moreover, the cognitive-neuropsychological approach to developmental disorders has been criticized on the grounds of the dynamics of the developing brain (Thomas and Karmiloff-Smith, 2002). Early acquired lesions or genetic dysfunctions can induce varying degrees of reorganization in the cognitive relevant brain processes. In exceptional cases, clear-cut structural-functional correlations, which are similar to the ones encountered in adults, are observed in cases of dysfunction in the developing brain (e.g., Temple, 1989, 1991; Sullivan et al., 1996; Ta'ir et al., 1997). In most cases of early acquired or genetic disorders, clinical-anatomical correlations are attenuated by several neuroplastic and compensatory processes.

Interestingly, there is an aspect of the performance of both H.V. and G.A. that could account for the patterns of the results observed in the respective cases without resorting to weak accounts that are based on typicality. The pattern of performance presented by H.V. reveals deficits in different numerical representations, which usually can be operated in an automatic or effortless fashion. The definition of the ANS, for example, involves an intuition for magnitudes and the capacity to activate it in a very automatic way (Dehaene, 1992; Verguts and Fias, 2008; Hyde, 2011). Moreover, the capacity to retrieve arithmetic facts appears to be a very automatic process as well (Domahs and Delazer, 2005; Zamarian et al., 2009). Such a specific deficit in the automatic access to information regarding, on the one hand, the ANS, and on the other hand, multiplication facts can account for the apparently discrepant deficits that are presented by H.V. A core deficit in the number sense alone cannot account for H.V.'s isolated deficits in multiplication but lack of deficit in subtraction operations of comparable difficulty.

On the other hand, the patterns of deficits presented by G.A. are suggestive of difficulties with a more executive and effortful processing of numerical representations as well as mit some aspects of effortless processing. The spared performance of G.A. in all arithmetic operations is compatible with this view because the problems employed in the present study never had operands that were larger than two-digits, with which G.A. has had sufficient experience in the past. In contrast, the transcoding task employed much larger numbers. This more complex part of the verbal numerical system is learned for the first time exactly in the grade that G.A. was attending during his assessment. This finding is suggestive that G.A. still needs substantial executive resources to employ correctly the transformation rules that are necessary to transcode those numbers (Barrouillet et al., 2004; Camos, 2008; Moura et al., 2013). More detailed analysis of G.A.'s poor transcoding performance reveals no evidence for a deficit in knowledge about the structure of the Portuguese verbal

number system. In contrast, G.A.'s error pattern is indicative of severe problems with phonological representations, concentration and monitoring capacity. Accordingly, orally presented word problems can also be more challenging for G.A. because a good capacity in verbal working memory is necessary to select relevant information from these problems and then operate with them until the correct result is obtained.

Support for this interpretation of H.V. and G.A. endophenotypes comes also from the analysis of the Five-digits Test results (see **Figure 2**). The Five-digits Test is well-suited to perform this comparison because the stimuli and task context are preserved, while the cognitive demands in terms of automatic and controlled processing vary (see also van der Sluis et al., 2004). On the one hand, H.V. presents difficulties with speeded counting but does not present difficulties with the inhibition- and shifting-demanding tasks. G.A., on the other hand, encounters difficulties in all aspects of the task, which requires both automatic and controlled processing. G.A.'s pattern of performance in the Five-digits Test is similar to the pattern observed by van der Sluis et al. (2004) on an equivalent numerical task in children with math learning difficulties and both math and reading learning difficulties. Interactions between processing speed and working memory impairments have been observed in several studies of both typically developing children (Berg, 2008) and children with math learning disability (Bull and Johnston, 1997). Moreover, disorders of automatization and procedural learning have also been implicated in learning disabilities of both reading (Menghini et al., 2006) and arithmetic (Lonnemann et al., 2011). Our results suggest that, in some cases, difficulties can be more related to the automatic or effortless processing, with possible compensation through more controlled strategies (H.V.), while in other cases, difficulties could be mixed or impairing more heavily controlled forms of processing (G.A.).

Overall, these results suggest that the search for endophenotypes could be more complex than originally expected, but not useless. In contrast, endophenotypes could be the only way to disclose more precise details on the nature and extension of mathematics difficulties. The current models of mathematics difficulties (e.g., Rubinsten and Henik, 2009) treat the different subtypes of math difficulties as members of a class of disorders that have different natures, which are nevertheless at more or less the same hierarchical level of organization of the cognitive system. This model has been proven to be useful but requires better specification.

One might consider the role of good executive functioning resources as a compensatory mechanism in developmental disorders. Johnson (2012) has proposed a role for executive functions in compensating for developmental neurogenetic impairments. According to this view, impairments in more basic and modularly organized aspects of information processing, such as phonological processing and number sense, can be compensated for if they are not sufficiently severe or if the individual has good executive functioning resources. The expression of symptoms that lead to diagnosis would occur in cases in which specific processing deficits are severe or when executive functioning resources are not sufficient to meet the environmental demands. The pattern of deficits presented by H.V. and G.A. are in line with these

arguments. While H.V. was able to mobilize resources from executive functions and compensate for many of her deficits in number processing, the same could not be observed in the case of G.A.

Moreover, H.V.'s case also suggests that, in addition to executive functions, a more basic level of task automatization should be considered to be a bridge between domain-specific and domain-general cognitive impairments that contribute to math learning difficulties. This topic has received less consideration in the literature (however, see van der Sluis et al., 2004; Chan and Ho, 2010).

Automatic and controlled processing are two dimensions of cognitive abilities that interact with domain-general and -specific factors, and the neurobiological basis of these processes should also be examined in more detail. Contemporary models of skill learning and automatization assume that, in the initial steps of learning, higher demands on processing are imposed over the fronto-parietal circuits that underlie cognitive control (Schneider and Chein, 2003). With practice, the typical focus of activity is shifted from anterior cortical regions to posterior ones and to the striatum. Another assumption is that this anterior-to-posterior shift in activity is domain-general because this circumstance has been observed with several motor and cognitive tasks. The extant literature largely supports these assertions (Patel et al., 2012). Similar observations have been made in the domain of numerical cognition. Interference effects in a number-size interference task are related to activation in frontal areas, while the distance effect is associated with activation in parietal areas, including the intraparietal sulcus (Kaufmann et al., 2005). Learning arithmetic facts is followed by a shift of the activation focus from frontal and intraparietal areas to the left angular gyrus (Zamarian et al., 2009). Developmentally, children usually activate more widespread areas during mental calculations, including frontal regions (Kawashima et al., 2004; Rivera et al., 2005). In adults, the focus of activity is more concentrated on posterior areas (Kaufmann et al., 2008, 2011; Klein et al., 2009).

Available evidence on the neurocognitive underpinnings of skill learning and automatization allow us to tentatively predict structural-anatomical correlations of automatic and controlled processing impairments in math learning difficulties. Numerical-specific automatic processing deficits, such as the deficits presented by H.V., should be related to impairments in parietal areas, including connections to the intraparietal sulcus. A broader pattern of dysfunction, encompassing the frontal areas, should be observed in cases such as G.A., in whom controlled processing is also impaired. Obviously, math learning difficulties that are associated with dyslexia also imply malfunctioning of perisylvian areas.

Results of the present paper have important implications for future research. The first implication is the need to include both domain-specific and domain-general measures to fully describe the range of manifestations and impairments in math learning difficulties (Cowan and Powell, 2013). Moreover, the neuropsychological test batteries that are used to assess math learning should fairly measure both automatic or effortless processing and effortful or controlled processing. Because comorbidity with ADHD can explain impairments in working memory and executive functions, ADHD symptoms should also necessarily

be controlled. Otherwise, it is not possible to draw straightforward conclusions on how working memory/executive functions determine more general performance difficulties compared to numerical-specific deficits (Willburger et al., 2008). Tasks that assess working memory and executive functions should also be presented in two formats, using non-numerical and numerical stimuli (Raghubar et al., 2010). Another important implication is the need to assess more automatized number processing, such as RAN. Finally, we believe that the present study contributes to underline the importance of single-case research in clarifying the role of distinct endophenotypes in dyscalculia research.

In the present study, we demonstrated that automatic and controlled information processing is one valid and necessary axis of investigation when characterizing the multitude of cognitive deficits that are associated with math difficulties, which can conciliate apparent discrepancies between individual and typical endophenotypes with respect to math difficulties. This approach constitutes a more general level of description of cognitive deficits as that originally adopted by other authors in previous studies. In summary, phenotypic manifestations of learning disabilities are compounded by impairments in both specific and general information processing mechanisms. Math-specific factors, such as number sense, and math-nonspecific cognitive factors, such as phonological processing, interact with general aspects of information processing, such as controlled processing and automatization. Math-specific and more general information processing deficits and automatic and controlled information processing deficits therefore represent orthogonal but interacting dimensions of the same disorder. In this sense, symptoms would be apparent when general or specific compensatory mechanisms are overloaded or not sufficient to meet the environmental demands in cases of more specific impairments. Impairments in restricted, specific domains could explain the unique difficulties, while impairment in more general mechanisms could be related to the degree and form of phenotypic expression via compensatory mechanisms.

ACKNOWLEDGMENTS

REFERENCES

Achenbach, T. M., Becker, A., Döpfner, M., Heiervang, E., Roessner, V., Steinhausen, H. C., et al. (2008). Multicultural assessment of children and adolescent psychopathology with ASEBA and SDQ instruments: research findings, applications, and future directions. *J. Child. Psychol. Psychiatry* 49, 251–275. doi: 10.1111/j.1469-7610.2007.01867.x

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders. DSM-IV-TR, 4th Edn., Text revision*. Washington, DC: American Psychiatric Press.

Angelini, A. L., Alves, I. C. B., Custódio, E. M., Duarte, W. F., and Duarte, J. L. M. (1999). *Matrizes Progressivas Coloridas de Raven - Escala Especial*. São Paulo: Centro Editor de Testes e Pesquisas em Psicologia.

Barrouillet, P., Camos, V., Perruchet, P., and Seron, X. (2004). ADAPT: a developmental, asemantic, and procedural model for transcoding from verbal to arabic numerals. *Psychol. Rev*. 111, 368–394. doi: 10.1037/0033-295X.111.2.368

Batstra, L., Neeleman, J., and Hadders-Algra, M. (2003). The neurology of learning and behavioral problems in pre-adolescent children. *Acta Psychiatr. Scand*. 108, 92–100. doi: 10.1034/j.1600-0447.2003.00127.x

Benn, Y., Zheng, Y., Wilkinson, I. D., Siegal, M., and Varley, R. (2012). Language in calculation: a core mechanism? *Neuropsychologia* 50, 1–10. doi: 10.1016/j.neuropsychologia.2011.09.045

Berg, D. H. (2008). Working memory and arithmetic calculation in children: the contributory roles of processing speed, short-term memory, and reading. *J. Exp. Child Psychol*. 99, 288–308. doi: 10.1016/j.jecp.2007.12.002

Birnboim, S. (2003). The automatic and controlled information-processing dissociation: is it still relevant? *Neuropsychol. Rev*. 13, 19–31. doi: 10.1023/A:1022348506064

Bishop, D. V. M., and Rutter, M. (2009). "Neurodevelopmenal disorders: conceptual issues," in *Rutter's Child and Adolescent Psychiatry, 5th Edn.*, eds M. Rutter, D. V. M. Bishopt, D. S. Pine, S. Scott, J. Stevenson, E. Taylor, and A. Thapar (Oxford: Blacwell), 32–41.

Bull, R. S., and Johnston, R. S. (1997). Children's arithmetical difficulties: contributions from processing speed, item identification, and short-term memory. *J. Exp. Child Psychol*. 65, 1–14. doi: 10.1006/jecp.1996.2358

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

Butterworth, B., and Reigosa, V. (2007). "Information processing deficits in dyscalculia," in *Why is Math so Hard for Some Children? The Nature and Origins of Mathematical Learning Difficulties and Disabilities*, eds D. B. Berch and M. M. M. Mazzocco (Baltimore, MD: Brookes), 65–81.

Camos, V. (2008). Low working memory capacity impedes both efficiency and learning of number transcoding in children. *J. Expl. Child Psychol*. 99, 37–57. doi: 10.1016/j.jecp.2007.06.006

Castles, A., and Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition* 91, 77–111. doi: 10.1016/S0010-0277(03)00164-1

Chan, B. M., and Ho, C. S. (2010). The cognitive profile of Chinese children with mathematics difficulties. *J. Exp. Child Psychol*. 107, 260–279. doi: 10.1016/j.jecp.2010.04.016

Charchat-Fichman, H., and Oliveira, R. M. (2009). Performance of 119 Brazilian children in the stroop paradigm - victoria version. *Arq. Neuropsiquiatr*. 67, 445–449. doi: 10.1590/S0004-282X2009000300014

Costa, A. J., Silva, J. B. L., Chagas, P. P., Krinzinger, H., Lonneman, J., Willmes, K., et al. (2011). A hand full of numbers: a role for offloading in arithmetics learning? *Front. Psychol*. 2:368. doi: 10.3389/fpsyg.2011.00368

Cowan, R., and Powell, D. (2013). The contributions of domain-general and numerical factors to third-grade arithmetic skills and mathematical learning disability. *J. Educ. Psychol*. 106, 214–229. doi: 10.1037/a0034097

Crawford, J. R., Garthwaite, P. H., and Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: rationale, methods, implementations, and proposed reporting standards. *Cogn. Neuropsychol*. 27, 245–260. doi: 10.1080/02643294.2010.513967

Deary, I. J., Strand, S., Smith, P., and Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence* 35, 13–21. doi: 10.1016/j.intell.2006.02.001

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N

Dehaene, S., and Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Math. Cogn*. 1, 83–120.

Dehaene, S., Izard, V., and Piazza, M. (2005). Control over non-numerical parameters in numerosity experiments. Available online at: http://www.unicog.org/docs/DocumentationDotsGeneration.doc

de Jong, C. G. W., Oosterlaan, J., and Sergeant, J. A. (2006). The role of double dissociation studies in the search for candidate endophenotypes for the comorbidity of attention deficit hyperactiivity disorder and reading disability. *Int. J. Disabil. Dev. Educ*. 53, 177–193. doi: 10.1080/10349120600716158

de Jong, C. G. W., van de Voorde, S., Roeyers, H., Raymaekers, R., Oosterlaan, J., and Sergeant, J. A. (2009). How distinctive are ADHD and RD? Results of a double dissociation study. *J. Abnorm. Child Psychol.* 37, 1007–1017. doi: 10.1007/s10802-009-9328-y

Dellatolas, G., Viguier D., Deloche G., and De Agostini, M. (1998). Right-left orientation and significance of systematic reversal in children. *Cortex* 34, 659–676. doi: 10.1016/S0010-9452(08)70771-8

Denckla, M. B. (1997). "The neurobehavioral examination in children," in *Behavioral Neurology and Neuropsychology*, eds T. E. Feinberg and M. J. Farah (New York, NY: McGraw-Hill), 721–728.

Denckla, M. B. (2003). ADHD: topic update. *Brain Dev.* 25, 383–389. doi: 10.1016/S0387-7604(03)00057-3

De Smedt, B., and Boets, B. (2010). Phonological processing and arithmetic fact retrieval: evidence from developmental dyslexia. *Neuropsychologia* 48, 3973–3981. doi: 10.1016/j.neuropsychologia.2010.10.018

De Smedt, B., Taylor, J., Archibald, L., and Ansari, D. (2010). How is phonological processing related to individual differences in children's arithmetic skills? *Dev. Sci.* 13, 508–520. doi: 10.1111/j.1467-7687.2009.00897.x

de Visscher, A., and Noël, M. P. (2013). A case study of arithmetic facts dyscalculia caused by a hypersensitivity-to-interference in memory. *Cortex* 49, 50–70. doi: 10.1016/j.cortex.2012.01.003

Domahs, F., and Delazer, M. (2005). Some assumptions and facts about arithmetic facts. *Psychol. Sci.* 47, 96–111.

Ferreira, F. O., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., et al. (2012). Explaining school mathematics performance from symbolic and nonsymbolic magnitude processing: similarities and differences between typical and low-achieving children. *Psychol. Neurosci.* 5, 37–46. doi: 10.3922/j.psns.2012.1.06

Figueiredo, V. L. M. (2002). *WISC-III: Escala de Inteligência Wechsler para Crianças. Manual Adaptação e Padronização Brasileira.* São Paulo: Casa do Psicólogo.

Figueiredo, V. L. M., and do Nascimento, E. (2007). Desempenhos nas duas tarefas do subteste Dígitos do WISC-III e do WAIS-III. *Psicol. Teor. Pesq.* 23, 313–138. doi: 10.1590/S0102-37722007000300010

Fine, J. G., Semrud-Clikeman, M., Bledsoe, J. C., and Musielak, K. A. (2013). A critical review of the literature on NLD as a developmental disorder. *Child Neuropsychol.* 19, 90–223. doi: 10.1080/09297049.2011.648923

Geary, D. C. (1993). Mathematical disabilities: cognitive, neuropsychological, and genetic components. *Psychol. Bull.* 114, 345–362. doi: 10.1037/0033-2909.114.2.345

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., and Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Dev.* 78, 1343–1359. doi: 10.1111/j.1467-8624.2007.01069.x

Gross-Tsur, V., Manor, O., and Shalev, R. S. (1996). Developmental dyscalculia: prevalence and demographic features. *Dev. Med. Child Neurol.* 38, 25–33. doi: 10.1111/j.1469-8749.1996.tb15029.x

Halberda, J., Mazzocco, M. M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* 455, 665–669. doi: 10.1038/nature07246

Hasher, L., and Zacks, R. T. (1979). Automatic and effortful processes in memory. *J. Exp. Psychol.* 108, 356–388. doi: 10.1037/0096-3445.108.3.356

Hecht, S. A., Torgesen, J. K., Wagner, R. K., and Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: a longitudinal study from second to fifth grades. *J. Exp. Child Psychol.* 79, 192–227. doi: 10.1006/jecp.2000.2586

Huber, P. J. (1973). *Psychometrische Einzelfalldiagnostik*. Beltz: Weinheim.

Hulme, C., Bowyer-Crane, C., Carroll, J., Duff, F., and Snowling, M. (2012). The causal role of 2 phoneme awareness and letter-sound knowledge in learning to read. *Psychol. Sci.* 23, 572–577. doi: 10.1177/0956797611435921

Hyde, D. C. (2011). Two systems of non-symbolic numerical cognition. *Front. Hum. Neurosci.* 5:150. doi: 10.3389/fnhum.2011.00150

Izard, V., and Dehaene, S. (2008). Calibrating the mental number line. *Cognition* 106, 1221–1247. doi: 10.1016/j.cognition.2007.06.004

Johnson, M. H. (2012). Executive function and developmental disorders: the flip side of the coin. *Trends Cogn. Sci.* 16, 454–457. doi: 10.1016/j.tics.2012.07.001

Júlio-Costa, A., Antunes, A. M., Lopes-Silva, J. B., Moreira, B. C., Vianna, G. S., Wood, G., et al. (2013) Count on dopamine: influences of COMT polymorphisms on numerical cognition. *Front. Psychol.* 4:531. doi: 10.3389/fpsyg.2013.00531

Kaufmann, L. (2002). More evidence for the role of the central executive in retrieving arithmetical facts - a case study of severe developmental dyscalculia. *J. Clin. Exp. Neuropsychol.* 24, 302–310. doi: 10.1076/jcen.24.3.302.976

Kaufmann, L., Lochy, A., Drexler, A., and Semenza, C. (2004). Deficient arithmetic fact retrieval - storage or access problem? A case study. *Neuropsychologia* 42, 482–496. doi: 10.1016/j.neuropsychologia.2003.09.004

Kaufmann, L., Koppelstaetter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., et al. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: an event-related fMRI study. *Neuroimage* 25, 888–898. doi: 10.1016/j.neuroimage.2004.12.041

Kaufmann, L., Vogel, S. E., Wood, G., Kremser, C., Schocke, M., Zimmerhackl, L.-B., et al. (2008). A developmental fMRI study of nonsymbolic numerical and spatial processing. *Cortex* 44, 376–385. doi: 10.1016/j.cortex.2007.08.003

Kaufmann, L., Wood, G., Rubinsten, O., and Henik, A. (2011). Meta-analyses of developmental fMRI studies investigating typical and atypical trajectories of number processing and calculation. *Dev. Neuropsychol.* 36, 763–787. doi: 10.1080/87565641.2010.549884

Kawashima, R., Taira, M., Okita, K., Inoue, K., Tajima, N., Yoshida, H., et al. (2004). A functional MRI study of simple arithmetic—a comparison between children and adults. *Brain Res. Cogn. Brain Res.* 18, 227–233. doi: 10.1016/j.cogbrainres.2003.10.009

Klein, E., Moeller, K., Dressel, K., Domahs, F., Wood, G., Willmes, K., et al. (2010). To carry or not to carry–is this the question? Disentangling the carry effect in multi-digit addition. *Acta Psychol.* 135, 67–76. doi: 10.1016/j.actpsy.2010.06.002

Klein, E., Nuerk, H. C., Wood, G., Knops, A., and Willmes, K. (2009). The exact vs. approximate distinction in numerical cognition may not be exact, but only approximate: How different processes work together in multi-digit addition. *Brain Cogn.* 69, 369–381. doi: 10.1016/j.bandc.2008.08.031

Landerl, K., Fussenegger, V., Moll, K., and Willburger, E. (2009). Dyslexia and dyscalculia: two learning disorders with different cognitive profiles. *J. Exp. Child Psychol.* 103, 309–324. doi: 10.1016/j.jecp.2009.03.006

Landerl, K., and Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. *J. Exp. Child Psychol.* 103, 546–565. doi: 10.1016/j.jecp.2008.12.006

Landerl, K., and Moll, K. J. (2010). Comorbidity of learning disorders: prevalence and familial transmission. *J. Child Psychol. Psychiatry* 51, 287–294. doi: 10.1111/j.1469-7610.2009.02164.x

Lefèvre, A. B., and Diament, A. J. (1982). Epidemiologia em neurologia infantil: estudo dos diagnósticos mais comuns. *Rev. Hosp. Clin. Fac. Med. Sao Paulo*, 37, 199–205.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychol. Rev.* 95, 492. doi: 10.1037/0033-295X.95.4.492

Lonnemann, J., Linkersdörfer, J., Hesselhaus, V., Hasselhorn, M., and Lindberg, S. (2011). Relations between balancing and arithmetic skills in children - evidence of cerebellar impairment? *J. Neurolinguistics* 24, 592–601. doi: 10.1016/j.jneuroling.2011.02.005

Lopes-Silva, J. B., Moura, R., Júlio-Costa, A., Haase, V. G., Wood, G. (2014) Phonemic awareness as a pathway to number transcoding. *Fron. Psychol.* 5:13. doi: 10.3389/fpsyg.2014.00013

Mazzocco, M. M. M. (2007). "Defining and differentiating mathematical learning disabilities and difficulties," *Why is Math so Hard for Some Children? The Nature and Origins of Mathematical Learning Difficulties and Disabilities*, eds D. B. Berch and M. M. M. Mazzocco (Baltimore, MD: Brookes), 29–47.

Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011a). Preschooler's precision of the approximate number system predicts later school mathematics performance. *PLoS ONE* 6:e23749. doi: 10.1371/journal.pone.0023749

Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011b). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Dev.* 82, 1224–1237. doi: 10.1111/j.1467-8624.2011.01608.x

McCloskey, M., Caramazza, A., and Basili, A. (1985). Cognitive mechanisms in number processing and calculation: evidence from dyscalculia. *Brain Cogn.* 4, 171–196. doi: 10.1016/0278-2626(85)90069-7

Melby-Lervåg, M., Lyster, S., and Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *Psychol. Bull.* 138, 322–352. doi: 10.1037/a0026744

Menghini, D., Hagberg, G. E., Caltagirone, C., Petrosini, L., and Vicari, S. (2006). Implicit learning deficits in dyslexic adults: an fMRI study. *Neuroimage* 33, 1218–1226. doi: 10.1016/j.neuroimage.2006.08.024

Moura, R., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., et al. (2013). Transcoding abilities in typical and atypical mathematics achievers: the role of working memory and procedural and lexical competencies. *J. Exp. Child Psychol.* 116, 707–727. doi: 10.1016/j.jecp.2013.07.008

Moyer, R. S., and Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature* 215, 1519–1520. doi: 10.1038/2151519a0

Noël, M. P., and Rousselle, L. (2011). Developmental changes in the profiles of dyscalculia: na explanation based on a double exact-and-approximate number representation model. *Front. Hum. Neurosci.* 5:165. doi: 10.3389/fnhum.2011.00165

Oliveira, M. S. (1999). *Figuras complexas de Rey: teste de cópia e de reprodução de memória de figuras geométricas complexas. Manual andré rey.* São Paulo, SP: Casa do Psicólogo.

Oliveira-Ferreira, F., Costa, D. S., Micheli, L. R., Pinheiro-Chagas, P., and Haase, V. G. (2012). School achievement test: normative data for a representative sample of elementary school children. *Psychol. Neurosci.* 5, 157–164. doi: 10.3922/j.psns.2012.2.05

Patel, R., Spreng, R. N., and Turner, G. R. (2012). Functional brain changes following cognitive and motor skills training: a quantitative meta-analysis. *Neurorehabil. Neural Repair* 27, 187–199. doi: 10.1177/1545968312461718

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends Cogn. Sci.* 14, 542–551. doi: 10.1016/j.tics.2010.09.008

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012

Pixner, S., Zuber, J., Hermanová, V., Kaufmann, L., Nuerk, H. C., and Moeller, K. (2011). One language, two number-word systems and many problems: numerical cognition in the Czech language. *Res. Dev. Disabil.* 32, 2683–2689. doi: 10.1016/j.ridd.2011.06.004

Poole, J. L., Burtner, P. A., Torres, T. A., McMullen, C. K., Markham, A., Marcum, M. L., et al. (2005). Measuring dexterity in children using the Nine-hole Peg Test. *J. Hand. Ther.* 18, 348–351. doi: 10.1197/j.jht.2005.04.003

Price, G. R., and Ansari, D. (2013). Dyscalculia, characteristics, causes, and treatments. *Numeracy* 6, 2. doi: 10.5038/1936-4660.6.1.2

Power, R. J. D., and Dal Martello, M. F. (1990). The dictation of Italian numerals. *Lang. Cogn. Process.* 5, 237–254. doi: 10.1080/01690969008402106

Punt, M., de Jong, M., de Groot, E., and Haaders-Algra, M. (2010). Minor neurological dysfunction in children with dyslexia. *Dev. Med. Child Neurol.* 52, 1127–1132. doi: 10.1111/j.1469-8749.2010.03712.x

Raghubar, K. P., Barnes, M. A., and Hect, S. A. (2010). Working memory and mathematics: a review of developmental, individual difference, and cognitive approaches. *Learn. Individ. Differ.* 20, 110–122. doi: 10.1016/j.lindif.2009.10.005

Reigosa-Crespo, V., Valdés-Sosa, M., Butterworth, B., Estévez, N., Rodríguez, M., Santos, E., et al. (2012). Basic numerical capacities and prevalence of developmental dyscalculia: the Havana survey. *Dev. Psychol.* 48, 123–145. doi: 10.1037/a0025356

Rivera, S. M., Reiss, A. L., Eckert, M. A., and Menon, V. (2005). Developmental changes in mental arithmetic: evidence for increased functional specialization in the left inferior parietal cortex. *Cereb. Cortex* 15, 1779–1790. doi: 10.1093/cercor/bhi055

Rocha, M. M., Rescorla, L. A., Emerich, D. R., Silvares, E. F. M., Borsa, J. C., Araújo, L. G. S., et al. (2012). Behavioural/emotional problems in Brazilian children: findings fromparents' reports on the Child Behavior Checklist. *Epidemiol. Psychiatr. Sci.* 22, 329–338. doi: 10.1017/S2045796012000637

Rousselle, L., and Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: a comparison of symbolic vs nonsymbolic number magnitude processing. *Cognition* 102, 361–395. doi: 10.1016/j.cognition.2006.01.005

Rubinsten, O., and Henik, A. (2005). Automatic activation of internal magnitudes: a study of developmental dyscalculia. *Neuropsychology* 19, 641. doi: 10.1037/0894-4105.19.5.641

Rubinsten, O., and Henik, A. (2006). Double dissociation of functions in developmental dyslexia and dyscalculia. *J. Educ. Psychol.* 98, 854. doi: 10.1037/0022-0663.98.4.854

Rubinsten, O., and Henik, A. (2009). Developmental dyscalculia: heterogeneity might not mean different mechanisms. *Trends Cogn. Sci.* 13, 92–99. doi: 10.1016/j.tics.2008.11.002

Santos, F. H., and Bueno, O. F. A. (2003). Validation of the Brazilian children's test of pseudoword repetition in Portuguese spears ages 4 to 10 years. *Braz. J. Med. Biol. Res.* 36, 1533–1547. doi: 10.1590/S0100-879X2003001100012

Santos, F. H., Mello, C. B., Bueno, O. F. A., and Dellatolas, G. (2005). Cross-cultural differences for three visual memory tasks in Brazilian children. *Percept. Mot. Skills* 101, 421–433. doi: 10.5209/rev_SJOP.2012.v15.n2.38862

Schneider, W., and Chein, J. M. (2003). Controlled & automatic processing: behavior, theory, and biological mechanisms. *Cogn. Sci.* 27, 525–559. doi: 10.1207/s15516709cog2703_8

Sedó, M. A. (2004). 5 digit test: a multilinguistic non-reading alternative to the Stroop test. *Rev. Neurol.* 38, 824–828.

Sedó, M. A. (2007). *FDT. Teste de los Cinco Dígitos. Manual.* Madrid: TEA.

Shiffrin, R. M., and Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychol. Rev.* 84, 127. doi: 10.1037/0033-295X.84.2.127

Stein, L. M. (1994). *Teste de Desempenho Escolar.* São Paulo: Casa do Psicólogo.

Strauss, E., Sherman, E. M. S., and Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary, 3rd Edn.* New York, NY: Oxford University Press.

Sullivan, K. S., Macaruso, P., and Sokol, S. M. (1996). Remediation of Arabic numeral processing in a case of developmental dyscalculia. *Neuropsycholl. Rehabil.* 6, 27–53. doi: 10.1080/713755495

Ta'ir, J., Brezner, A., and Ariel, R. (1997). Profound developmental dyscalculia: evidence for a cardinal/ordinal acquisition device. *Brain Cogn.* 35, 184–206. doi: 10.1006/brcg.1997.0937

Temple, C. M. (1989). Digit dyslexia: a category-specific disorder in developmental dyscalculia. *Cogn. Neuropsychol.* 6, 93–116. doi: 10.1080/02643298908253287

Temple, C. M. (1991). Procedural dyscalculia and number fact dyscalculia: double dissociation in developmental dyscalculia. *Cogn. Neuropsychol.* 8, 155–176. doi: 10.1080/02643299108253370

Temple, C. M. (1997). *Developmental Cognitive Neuropsychology.* Hove: Psychology Press.

Temple, C. M., and Clahsen, H. (2002). How connectionis simulations fail to account for developmental disorders in childern. *Behav. and Brain Sci.* 25, 769–770. doi: 10.1017/S0140525X02410138

Temple, C. M., and Sherwood, S. (2007). Representation and retrieval of arithmetic facts: developmental difficulties. *Q. J. Exp. Psychol.* 55, 733–752. doi: 10.1080/02724980143000550

Thomas, M., and Karmiloff-Smith, A. (2002). Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. *Behav. Brain Sci.* 25, 727–788. doi: 10.1017/S0140525X02000134

Tressoldi, P., Rosati, M., and Lucangeli, D. (2007). Patterns of developmental dyscalculia with or without dyslexia. *Neurocase* 13, 217–225. doi: 10.1080/13554790701533746

van der Sluis, S., de Jong, P. F., and van der Leij, A. (2004). Inhibition and shifting in children with learning deficits in arithmetic and reading. *J. Exp. Child Psychol.* 87, 239–266. doi: 10.1016/j.jecp.2003.12.002

Varley, R. A., Klessinger, N. J. C., Romanowski, C. A. J., and Siegal, M. (2005). Agrammatic but numerate. *Proc. Natl. Acad. Sci. U.S.A.* 102, 3519–3524. doi: 10.1073/pnas.0407470102

Vaz, I. A., Cordeira, P. M., de Macedo, E. C., and Kukasova, K. (2010). Memória de trabalho em crianças avaliada pela tarefa de Brown-Peterson. *Pró-Fono.* 22, 95–100. doi: 10.1590/S0104-56872010000200005

Venneri, A., Cornoldi, C., and Garuti, M. (2003). Arithmetic difficulties in children with viusospatial learning disability (VLD). *Child Neuropsychol.* 9, 175–183. doi: 10.1076/chin.9.3.175.16454

Verguts, T., and Fias, W. (2008). Symbolic and nonsymbolic pathways of number processing. *Phil. Psych.* 21, 539–554. doi: 10.1080/09515080802285545

Wagner, R. K., and Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychol. Bull.* 101, 192–212. doi: 10.1037//0033-2909.101.2.192

White, S., Frith, U., Milne, E., Rosen, S., Swettenham, J., and Ramus, F. (2006a). A double dissociation between sensorimotor impairments and reading disability: a comparison of autistic and dyslexic children. *Cogn. Neuropsychol.* 23, 748–761. doi: 10.1080/02643290500438607

White, S., Milne, E., Rosen, S., Hansen, P., Swettenham, J., Frith, U., et al. (2006b). The role of sensorimotor impairments in dyslexia: a multiple case study of dyslexic children. *Dev. Sci.* 9, 237–269. doi: 10.1111/j.1467-7687.2006.00483.x

Willburger, E., Fussenegger, B., Moll, K., Wood, G., and Landerl, K. (2008). Naming speed in dyslexia and dyscalculia. *Learn. Individ. Differ.* 18, 224–236. doi: 10.1016/j.lindif.2008.01.003

Willmes, K. (1985). An approach to analyzing a single subject's scores obtained in a standardized test with application to the Aachen Aphasia Test (AAT). *J. Clin. Exp. Neuropsychol.* 7, 331–352. doi: 10.1080/01688638508401268

Willmes, K. (2003). "The methodological and statistical foundations of neuropsychological assessment," in *Handbook of Clinical Neuropsychology*, eds P. Halligan, U. Kischka, and J. Marshall (Oxford: Oxford University Press), 27–47.

Wilson, A. J., and Dehaene, S. (2007). "Number sense and developmental dyscalculia," in *Human Behavior, Learning, and the Developing Brain: A Typical Development*, eds D. Coch, G. Dawson, and K. W. Fischer (New York, NY: Guilford), 212–238.

World Health Organization. (2011). *ICD-10. International Statistical Classification of Diseases and Related Health Problems, 4th Edn.* Geneva: Author.

Zamarian, L., Ischebeck, A., and Delazer, M. (2009). Neuroscience of learning arithmetic–evidence from brain imaging studies. *Neurosci. Biobehav. Rev.* 33, 909–925. doi: 10.1016/j.neubiorev.2009.03.005

Zheng, X., Swanson, H. L., and Marcoulides, G. A. (2011). Working memory components as predictors of children's mathematical word problem solving. *J. Exp. Child Psychol.* 110, 481–498. doi: 10.1016/j.jecp.2011.06.001

The added value of eye-tracking in diagnosing dyscalculia: a case study

Sietske van Viersen, Esther M. Slot, Evelyn H. Kroesbergen *, Jaccoline E. van't Noordende and Paul P. M. Leseman

Department of Cognitive and Motor Disabilities, Utrecht University, Utrecht, Netherlands

The present study compared eye movements and performance of a 9-year-old girl with Developmental Dyscalculia (DD) on a series of number line tasks to those of a group of typically developing (TD) children ($n = 10$), in order to answer the question whether eye-tracking data from number line estimation tasks can be a useful tool to discriminate between TD children and children with a number processing deficit. Quantitative results indicated that the child with dyscalculia performed worse on all symbolic number line tasks compared to the control group, indicated by a low linear fit (R^2) and a low accuracy measured by mean percent absolute error. In contrast to the control group, her magnitude representations seemed to be better represented by a logarithmic than a linear fit. Furthermore, qualitative analyses on the data of the child with dyscalculia revealed more unidentifiable fixation patterns in the processing of multi-digit numbers and more dysfunctional estimation strategy use in one third of the estimation trials as opposed to ~10% in the control group. In line with her dyscalculia diagnosis, these results confirm the difficulties with spatially representing and manipulating numerosities on a number line, resulting in inflexible and inadequate estimation or processing strategies. It can be concluded from this case study that eye-tracking data can be used to discern different number processing and estimation strategies in TD children and children with a number processing deficit. Hence, eye-tracking data in combination with number line estimation tasks might be a valuable and promising addition to current diagnostic measures.

Keywords: dyscalculia, number sense, eye-tracking, number line, mapping, diagnostic procedures

The present study investigated the added value of number line estimation tasks combined with eye-tracking data for the diagnostic procedure of dyscalculia. Although little is known about the validity or utility of eye movements as a diagnostic measure of mathematical difficulties, eye-tracking seems a valid instrument to examine the underlying processes in number line estimation (Schneider et al., 2008; Heine et al., 2010; Sullivan et al., 2011). The use of eye movements in combination with number line tasks as a diagnostic instrument might provide important insights in the processing of (non-)numerical information and possible underlying deficits in children with dyscalculia. Current diagnostic measures are mainly focused on the mathematical achievement level and therefore less adequate for identifying a cognitive deficit in numerical processing skills. The eye-tracking data can reveal underlying processes and could also lead toward more knowledge about the use or absence of number processing and estimation strategies and their detectability in children with dyscalculia.

Developmental Dyscalculia (DD) is defined as a learning disability characterized by severe mathematical problems. Children with DD often show difficulties in representing and manipulating numerical information non-verbally and spatially, in the automation of arithmetic facts, and in executing arithmetic procedures (prevalence 3–6%; Rotzer et al., 2009;

De Visscher and Noel, 2012). Concretely, children with DD often display difficulties with for example approximation, counting sequences, and subitizing (e.g., Butterworth et al., 2011; Mazzocco et al., 2011; Desoete et al., 2012; Geary et al., 2012). DD has been linked to a core deficit in number sense, i.e., the spatial processing of (number) magnitude information (Dehaene, 1997; Wilson and Dehaene, 2007), possibly caused by impaired functioning of the intra parietal sulcus (IPS; Kucian et al., 2006; Rotzer et al., 2008). According to Dehaene (2001), number sense refers to the "fundamental ability to mentally represent and manipulate numerosities on a mental 'number line'" (p. 17), a line along which numerical magnitudes are represented in ascending order with respect to their magnitude (Moeller et al., 2009b). The development of the mental number line is based on the ability to map number symbols onto their corresponding non-symbolic magnitude (Kolkman et al., 2013). Between the ages of 3.5 and 8 years, typical children become increasingly aware of the concept of numerosity, i.e., the understanding that number symbols are connected to quantities (Kolkman et al., 2013). This change results from transcoding processes, enabling the child to translate information from non-symbolic to symbolic format and vice versa, providing the number symbols with a non-symbolic magnitude meaning (Dehaene, 2001). In line with Kolkman et al. (2013) we refer to these skills as "mapping skills."

Number line tasks are often used to measure mapping skills (e.g., Schneider et al., 2008; Kolkman et al., 2013). In these tasks, children have to estimate the position of a given number on a horizontal number line from 0 to 10, 0 to 100, or 0 to 1000. Previous research has shown that many children with DD have difficulties with estimating the correct positions of numbers on a number line (e.g., Kucian et al., 2006), probably caused by a deficient development of the mapping skills. Young children that have not fully developed mapping skills also tend to make inaccurate logarithmic placements instead of linear placements on a number line (Booth and Siegler, 2006). For example, Moeller et al. (2009b) found that young children tend to overestimate single-digit numbers, using up to half of the number line and placing all two-digit numbers in the remaining space. Larger and less familiar numbers also lead to a decrease in accuracy (Ebersbach et al., 2008). This indicates that the development of the mental number line in young children (kindergarten age) is typically characterized by a logarithmic representation, meaning that the distance between the magnitudes of numbers at the low end of the range is exaggerated and the distance between magnitudes of numbers in the middle and upper ends of the range are minimized (Siegler and Booth, 2004; Kucian et al., 2006). When number representations become more precise, a shift from a logarithmic to a linear ruler representation has been observed in typically developing (TD) children. The timing of this shift depends on the number range of the number line; the linear representation is observed earlier in small number ranges than in larger number ranges (Siegler and Booth, 2004; Booth and Siegler, 2006; Opfer and Siegler, 2007; Siegler et al., 2009; Friso-van den Bos et al., under review). Empirical evidence suggests that 10-year-old children with dyscalculia display number representations at the level of 5-year-old TD children (Piazza et al., 2010), although there is also evidence showing higher linear than logarithmic fit in children with math problems (e.g., Van't Noordende and Kolkman, 2013).

Concerning the performance of children with DD on number line tasks, it is important to note that a deficit in number sense is not the only factor that might explain poor results on the estimation of numbers. Number line tasks require specific working memory abilities such as visuospatial processing and visuospatial storage (Kolkman et al., under review). Children with DD often show working memory deficits, especially in the visuospatial domain (Kroesbergen et al., 2007; Toll et al., 2011; Geary et al., 2012; Passolunghi and Mammarella, 2012; De Weerdt et al., 2013). Poor spatial working memory processes might partly underlie the mapping deficit, inhibiting the formation of a mental number line as well as the storage and retrieval of arithmetic facts (Rotzer et al., 2009; Kolkman et al., under review). Consequently, working memory ability was included in this study as a background variable, as it may form an additional explanation for the low performance and strategy use of the child with DD on the number line task.

NUMBER PROCESSING AND ESTIMATION STRATEGIES

Several theories have been derived from eye-tracking research on the processing and estimation of (non-)symbolic magnitudes and single- and multiple-digit numbers. These theories describe observable strategies, indicated by eye movement patterns, in children with typically developing math skills. The present study will examine whether a child with dyscalculia displays atypical number processing and estimation strategies, which would confirm a deficit in (non-)numerical processing skills.

Research on numerical processing pertains solely to the processing of symbolic numbers, represented in Arabic symbols, consisting of two or more digits. Nuerk et al. (2011) provide a definition of multi-digit number processing, stating that processing of multiple-digit numbers depends on the integration or computation of at least two digits to realize a numerical entity. Roughly three strategies can be identified: (a) *holistic* processing, (b) *decomposed parallel* processing, and (c) *decomposed sequential* processing. According to Dehaene et al. (1990), a holistic strategy suggests that, before mental manipulation of numerical information takes place, magnitudes are transformed into distance measures mapped onto a mental number line. Numbers are assumed to be processed as an integrated entity, and not decomposed into tens and units (Dehaene et al., 1990). In contrast, in a broad review on multi-digit processing Nuerk et al. (2011) found evidence that the processing of two-digit numbers always occurs in a decomposed way. Decomposed processing means that numbers are "decomposed" into hundreds, tens, and units. However, there is little evidence about how decomposed processing functions. Poltrock and Schwartz (1984) argued for sequential decomposition, where multi-digit numbers are compared in a sequential digit-by-digit way, starting at the leftmost digit. In contrast, Nuerk and Willmes (2005) argue for parallel decomposition, implying that multi-digit numbers are decomposed and processed in a parallel fashion, considering the value of every digit in relation to the other.

All three strategies can be identified by specific eye movement patterns. As described by Moeller et al. (2009a), holistic processing can be identified by equal fixations at all digits. This way numbers are processed as an integrated entity, resulting in a "pictorial representation" of the number. Decomposed parallel processing is indicated by fixations at the largest entity first (e.g., tens) and less or shorter fixations at the smaller entity (e.g., units), carefully comparing the values of the digits from largest to smallest. Fixations only at the largest entity and (almost) no fixation at the smaller entity, or solely after the child has already started estimating, is indicative of decomposed sequential processing (Moeller et al., 2009a). So far, evidence was found for decomposed parallel processing of two-digit numbers (Moeller et al., 2009a) and three-digit numbers (Korvorst and Damian, 2008) in a general population. To the best of our knowledge, no evidence was found for the holistic and decomposed sequential strategies in eye-tracking data (Moeller et al., 2009a).

Research on estimation of magnitudes and numbers identified several strategies that can be used to estimate numbers on a line. Newman and Berger (1984) and Petitto (1990) stated that children in kindergarten and the first years of primary school were prone to use a *counting-up* or *counting-down* strategy; they start at one end of the line and count up or down in whole units or decades until they reach the target position. Older TD children tend to use a *midpoint* strategy; if the target position is closer to the midpoint of the line than to one of its ends, counting starts at the middle and subsequently up or down from there (Newman

and Berger, 1984; Petitto, 1990; Schneider et al., 2008). Ultimately, the use of these strategies is adapted to the target number that has to be estimated (Newman and Berger, 1984). Sullivan et al. (2011) showed that eye-tracking is a valid instrument to measure the use of these strategies. Eye movement patterns have been shown to be also adequate for detecting differences in strategy use (e.g., Schneider et al., 2008; Van't Noordende and Kolkman, 2013). Van't Noordende and Kolkman (2013) studied differences in eye movements between TD children and children with mathematical learning disabilities. They found that children with math learning disabilities use different strategies than children without math learning problems. They make less use of the reference points related to the specific strategies (i.e., begin, midpoint, and end). Moreover, most of their gazes to the reference points are on the midpoint and less on the begin- and endpoint of the line.

The present study compared the eye movements of a child with dyscalculia on a series of number line tasks to those of a control group in an explorative manner, in order to get a first indication of whether eye-tracking data from tasks tapping on number sense can be a useful tool in the diagnostic process of dyscalculia. It was hypothesized that the child with dyscalculia would show more atypical and more unidentifiable eye movement patterns than the control group when processing two-digit and three-digit numbers. In addition, it was hypothesized that the child would show more dysfunctional and more undefined strategies than the control group when estimating numbers on a number line, as she might make less use of reference points (i.e., begin, middle, end) and gaze relatively more at the midpoint (Van't Noordende and Kolkman, 2013). Moreover, it was expected that the eye-tracking data of the child with dyscalculia would indicate that her number representations were still logarithmic rather than linear, resulting in inaccuracy in estimation of multi-digit numbers as indicated by a higher mean percent absolute error than the control group. The data were analyzed using both quantitative and qualitative techniques.

METHODS

PARTICIPANTS

This case study concerned a 9-year-old Dutch girl (L), who had recently been diagnosed with DD at the diagnostic center of the Faculty of Social and Behavioral Sciences at Utrecht University (i.e., Ambulatorium), according to the three criteria as described by Van Luit et al. (2012). First, the diagnostic tests revealed that L's math abilities were significantly lower than expected based on her cognitive abilities (i.e., discrepancy criterion): L was tested as averagely intelligent, but on the timed math test she obtained a percentile score of 1. Second, L appeared to be substantially behind (i.e., more than 1.5 years) in basic numeracy skills and her level of automation (i.e., downfall criterion). Her numerical skills were comparable with the level of a student in second grade. Third, L did not benefit from the remedial teaching she received since Grade 2 (i.e., didactic resistance criterion).

The control group consisted of 10 TD children, of approximately the same age as L, from one school in the Netherlands ($M_{age} = 9.1$, $SD_{age} = 0.6$). All children performed within the average range on math ability (between the 50 and 75th percentile), as measured by a standardized Dutch complex

arithmetic test, and did not have a history of behavioral or learning problems.

INSTRUMENTS

Both the control group and L were assessed on a small range of behavioral and cognitive measures in order to compare their number sense, arithmetic and working memory abilities.

Mathematics

Recent percentile scores from a complex, standardized school arithmetic test (Janssen et al., 2005) were provided by the teachers of all control children to obtain information on their arithmetic abilities compared to the child with dyscalculia. The test consisted of 50 mathematical word problems and is generally administered at the end of the school year at almost all schools in The Netherlands.

Number sense

Three computerized number line tasks were used to measure symbolic and non-symbolic magnitude representations and multiple-digit representations. One non-symbolic number line task (i.e., 0–100) and two symbolic number line tasks (i.e., 0–100, and 0–1000) were included in the test battery, similar to the tasks used by Van't Noordende and Kolkman (2013). The outcomes of the tasks are R^2_{lin} and R^2_{log}, indicating the linear and logarithmic fit, which refer to the variance that is explained by a certain function. In addition, the mean percent absolute error for each child is computed, which is obtained by subtracting the correct quantity from the estimate and dividing it by the scale of estimates (e.g., 100 or 1000; Booth and Siegler, 2006).

Non-symbolic tasks. The non-symbolic number line task consisted of magnitudes in the form of dots, representing drops of fuel. The child was told to estimate how far the car could drive with a certain number of "fuel drops" and indicate this distance by placing the lever on the number line. The left side of the number line showed zero dots and the right side of the line showed 100 dots. The task consisted of 33 items.

Symbolic tasks. In the symbolic number line tasks, the child was asked to estimate magnitudes in the form of digits. The left side of the number line showed the number 0, the right side of the line showed to number 100 or 1000, and a random number was depicted beneath the number line in each trial. The child had to verbally repeat the number and then estimate were it would belong on the number line by placing the lever. The 0–100 and 0–1000 tasks both included 33 items.

Working memory

L's (working) memory capacity was measured using several subtests of the *Automated Working Memory Assessment* battery (AWMA; Alloway, 2007). Control children were also assessed on two short-term memory subtests (non-word recall and dot matrix). All subtests were discontinued after three incorrect answers. Percentile scores were reported. Verbal short-term memory was measured using digit recall, in which the child recalled increasing series of digits, and non-word recall, in which the child

recalled increasing series of non-words. Verbal working memory was measured by the listening recall subtest, in which the child had to answer questions, and after answering a series of questions, recall the first word of each question. Test-retest reliabilities of these subtests are good (Alloway et al., 2009).

Visuospatial short-term memory was measured using the subtest dot matrix, in which the child was shown a 4×4 grid of empty white squares, with red dots appearing in one square at a time. The child had to reproduce the sequence of the dot by pointing out the same sequence in an empty grid. The subtest odd-one-out was used to measure visuospatial working memory. This subtest required the child to indicate in increasingly complex sequences which figure out of three was odd, and recall the odd figures in a matrix. Test-retest reliabilities are good (Alloway et al., 2009).

Number processing and estimation strategies

The Tobii T60 Eye-Tracker was used to register the eye fixations during the number line tasks. This eye-tracker was installed on a Windows computer, situated in the lab of the university. Recordings of the eye movements, as well as heat maps and gaze plots could be derived from the eye-tracker for the analysis of the fixation patterns.

PROCEDURE

For L, an intern of the Ambulatorium administered the behavioral and cognitive measures of the assessment battery. A PhD-student administered the number line tasks and the eye-tracking data. Data was gathered within 1 h. For the control group, graduate students visited a school to assess 15 control children, of which 10 were included in the control group. The math test was administered in a classroom setting, the cognitive and number line tasks were all administered individually in a quiet room somewhere in the school. Data was gathered within 1 h for every child.

ANALYSIS

The data was analyzed both quantitatively and qualitatively. The quantitative analyses focused on the comparison of the performance on the number line tasks and working memory performance between L and the control group. For the number line tasks, the level of performance was indicated by measures of logarithmic and linear fit and the mean percent absolute error. For working memory, the level of performance was indicated by percentiles on two different memory components. The aim of these analyses was to establish the representativeness of the control group and to provide an insight in the specific deficits that L displays. To investigate processing and estimation strategies, qualitative analyses were used to observe and code the eye fixation patterns in the eye-tracking videos. An attempt was made to discern differences in strategy use between L and the control group. Since the use of eye-tracking data in combination with number line tasks is still experimental, no predefined coding schemes were available (i.e., in terms of pixel settings or fixation duration in milliseconds). Consequently, the eye fixation patterns were explored based on descriptive information derived from the literature on possible strategy use in number processing and estimation. Coding was performed by two trained PhD-students

with experience in analyzing eye-tracking data. Cohen's kappa was computed to determine the inter rater reliability, which was 0.77 for the processing strategies on both symbolic tasks and 0.64 and 0.71 for the estimation strategies on the 0–100 and 0–1000 symbolic number line tasks.

First, the number processing strategies were coded based on the information from the studies of Nuerk et al. (2011) and Moeller et al. (2009a). For each trial, the sequences of eye movements on the given numbers were determined while the child attempted to estimate its place on the number line. As elaborated on before, each strategy can be identified based on a specific fixation pattern. Holistic processing was identified by equal fixation (as determined by an expanding dot indicating fixation location and duration, see **Figure 1** for the representative dot sizes per strategy) at all digits at the same time, before the child moved the lever to the number line. Decomposed parallel processing was indicated by fixations starting at the largest entity (e.g., hundreds) and less or shorter fixations progressing toward the smaller entities (e.g., tens), all before the child moved the lever to the number line. Decomposed sequential processing was indicated by a fixation only at the largest entity, which happened before the child moved the lever to the number line, and (almost) no fixation at the smaller entity, which might also occur while the child is already lugging the lever across the number line. If the eye movements showed unclear sequences that did not correspond to any of the predefined strategies, the trial was coded as *undefined*. When the eye movements indicated a pattern that seemed not random but rather atypical, and also did not fit any of the three strategies that were derived from the literature [see "unexpected observed fixation pattern" in Moeller et al. (2009a)], the trial was coded as *other*. Examples of fixation patterns belonging
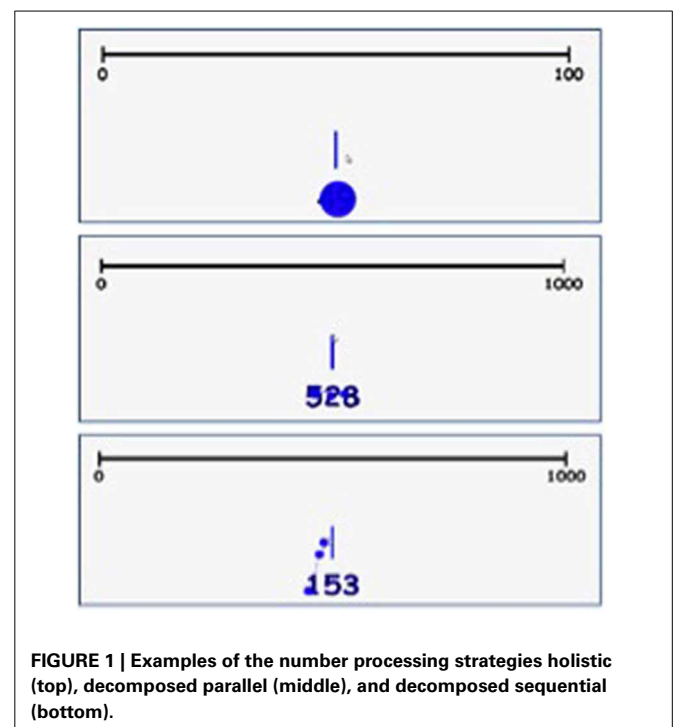
FIGURE 1 | Examples of the number processing strategies holistic (top), decomposed parallel (middle), and decomposed sequential (bottom).

to the different number processing strategies are displayed in **Figure 1**.

Second, the number estimation strategies were coded based on studies of Petitto (1990) and Newman and Berger (1984). Every strategy was identified based on the presence (or absence) of a specific fixation pattern. The *counting-up/counting-down* strategy indicated that the begin or the endpoint of the number line was used as reference to estimate a number. The *midpoint* strategy was identified by eye movements that indicated that the (exact) middle of the number line was used as a reference point (i.e., 50 in the 0–100 task and 500 in the 0–1000 task, with a margin of ±5% of the line length). In all three strategies, fixations at the chosen reference point had to be followed by the child step-by-step lugging the lever while estimating the given number. When a child showed fixation patterns that fitted multiple strategies or when it was not possible to discern a specific fixation pattern in the eye movements, the trial was coded as *undefined*. In addition to coding strategy use, it was determined per trial whether the used strategy was functional or dysfunctional. For making this distinction, the number line was divided in four equal quarters, which were each paired with a functional strategy. The *counting-up* strategy was considered functional for numbers between zero and 25(0), the *midpoint* strategy for numbers between 25(0) and 75(0), and the *counting-down* strategy for numbers between 75(0) and 100(0). Since this division seems rather arbitrary, we focused on relative differences between L and the control group in functional vs. dysfunctional strategy use, excluding the unidentifiable trials. Examples of fixation patterns belonging to the different number estimation strategies are displayed in **Figure 2**.

FIGURE 2 | Examples of the number estimation strategies counting-down/endpoint (top), midpoint (middle), and counting-up/beginpoint (bottom).
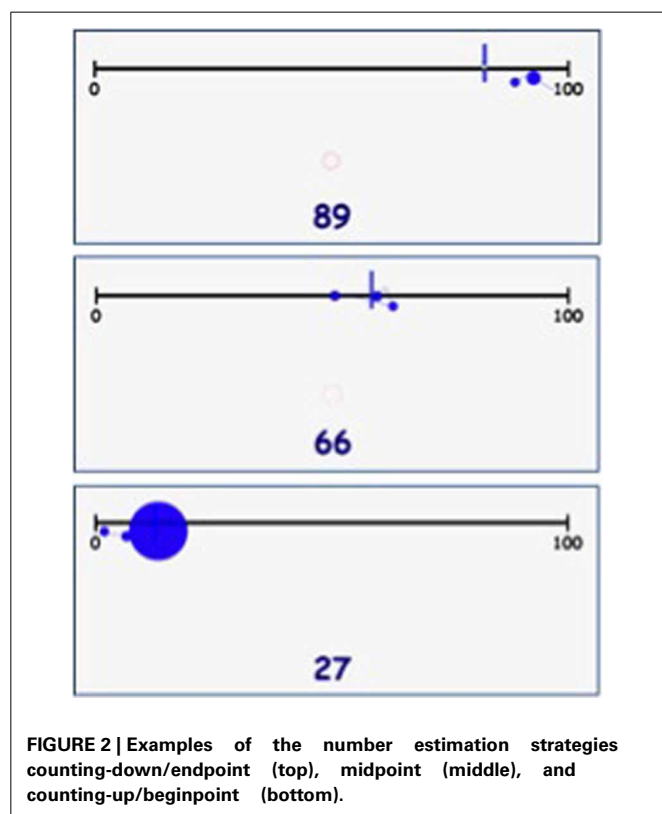
Differences between L and the control group were tested for significance with the SINGLIMS_ES.exe program, which is specifically designed for comparing a single case score to a control sample mean (Crawford and Howell, 1998; Crawford and Garthwaite, 2002; Crawford et al., 2010). Outcome measures are *p*-values and the z_{cc}-index, which represents an estimate of the average difference, measured in standard deviation units in order to be scale independent, between a case's score and the score of a randomly chosen member of the control population (Crawford et al., 2010).

RESULTS

QUANTITATIVE ANALYSES

The results of the (working) memory subtests displayed in **Table 1** show that L obtained scores below the 10th percentile on visuospatial short-term memory and visuospatial working memory. These scores indicate that L has a deficit in visuospatial abilities. The results indicate no deficit in verbal abilities. L obtained average scores on both the verbal working memory and verbal short-term memory subtests. The below average score on digit recall can be explained by the fact that this subtest involved numbers, which is her main weakness. In contrast to L's results, the control group scored within the average range (i.e., 25–75 percentile) on both visuospatial and verbal short-term memory, showing that the memory capacity of the control group is representative for the memory ability of the TD population.

The results on the number line tasks displayed in **Table 2** show that L's performance on the symbolic and the non-symbolic number line tasks was poor compared to the control group. On all symbolic tasks, her logarithmic and linear R^2-scores explained significantly less variance than in the control group, indicating that L has poor symbolic magnitude representations. Moreover, a comparison of the logarithmic and linear R^2 shows that L's number estimations on the non-symbolic 0–100 and symbolic 0–1000 number line tasks were better described by a logarithmic function than by a linear function, although the proportion of explained variance of L's $R^2_{\log}$-score on the 0–1000 number line is still considered very low (**Table 2**). In comparison, all linear R^2-scores of the control group showed a higher amount of

Table 1 | Working memory scores of the child with dyscalculia and the control group.

Scale	Subtest	Percentile
CHILD WITH DYSCALCULIA (*n* = 1)		
Verbal short-term memory	Non-word recall	22
	Digit recall	8
Verbal working memory	Listening recall	77
Visuospatial short-term memory	Dot matrix	1
Visuospatial working memory	Odd-one-out	9
CONTROL GROUP CHILDREN (*n* = 10)		
Verbal short-term memory	Non-word recall	48.1
Visuospatial short-term memory	Dot matrix	66.9

For the control group, the average percentile score per task is displayed.

explained variance, indicating good symbolic representations. In addition, the individual fit measures (see appendix) indicate that the number estimations of all children in the control group were better described by a linear function than a logarithmic function on the 0–1000 number line, which is considered age appropriate. As an illustration, the relations between the given numbers and the estimated numbers on the 0–1000 number line task of L and the control group, as represented by a linear and logarithmic regression line, are displayed in **Figure 3**. The R^2-scores on the symbolic tasks indicate that L's performance deteriorated more than the performance of the control group when magnitudes became larger. Although she already showed lower performance than the control group on the symbolic 0–100 number line, her performance on the 0–1000 number line was even worse, as indicated by a higher mean percent absolute error and lower linear fit.

A comparison of the mean percent absolute error between L and the control group shows that L's estimations were significantly more distanced from the given number than the control group's estimations (**Table 2**). All children in the control group showed a lower mean percent absolute error on both symbolic number line tasks than L. Hence, as expected, L's number line estimates were less accurate than the estimates of the TD children.

## QUALITATIVE ANALYSES
Processing strategies
For the qualitative analysis of multi-digit number processing, an attempt was made to discern specific fixation patterns from the eye movements on the symbolic 0–100 and 0–1000 tasks and connect these patterns to one of the processing strategies (i.e., holistic, decomposed parallel, and decomposed sequential). Trials were coded as *other* when a fixation pattern was observed but found atypical, as it could not be connected to a specific strategy, and *undefined* when no fixation pattern could be found in the eye movements. With L having severe and specific math problems, it was expected that she would show less strategy use and more atypical or unidentifiable sequences of eye fixations on the number line tasks compared to the control group.

On the symbolic 0–100 number line task, L showed equal fixations at both digits, indicating holistic processing, in 26.7% of the cases (i.e., 64, 53, 24, 49, 91, 43, 41, 99). In 13.3% of the cases (i.e., 96, 89, 87, 34) L fixated first at the tens and then briefly at the units, indicating decomposed parallel processing. Decomposed sequential processing was found in 10% of the cases (i.e., 72, 28, 60), as indicated by fixations at the tens and (almost) not at the units. In addition, also in 10% of the cases, L showed an atypical fixation pattern (i.e., 10, 37, 57) with eye movements starting at the smallest entity and then progressing toward the largest entity.

Table 2 | Summary of the results on the non-symbolic and symbolic number line tasks, number of trials, logarithmic and linear fit (R^2), and percent absolute error of the child with dyscalculia and the control group including *p*-values and effect sizes.

Number line	n	Child with dyscalculia (n = 1)			Control group (n = 10)			R^2_{lin}		R^2_{log}		Absolute error	
		R^2_{lin}	R^2_{log}	Absolute error (%)	R^2_{lin}	R^2_{log}	Absolute error (%)	*p*-value	z_{cc}	*p*-value	z_{cc}	*p*-value	z_{cc}
NON-SYMBOLIC													
0–100	33	0.699	0.712	17.7	0.783	0.695	13.0	.143	−1.19	.454	0.13	.147	1.17
SYMBOLIC													
0–100	33	0.849	0.698	10.7	0.955	0.780	4.3	.015	−2.70	.002	−4.10	.000	5.77
0–1000	33	0.268	0.325	20.1	0.930	0.602	5.8	.000	−22.07	.000	13.85	.000	13.75

n, number of trials, excluding two practice trials; R^2_{lin} and R^2_{log} represent the linear and logarithmic fit of the estimated numbers; Absolute error, mean percent absolute error. Z_{cc}, obvious direct analog of Cohen's d, subscript representing "case-controls."
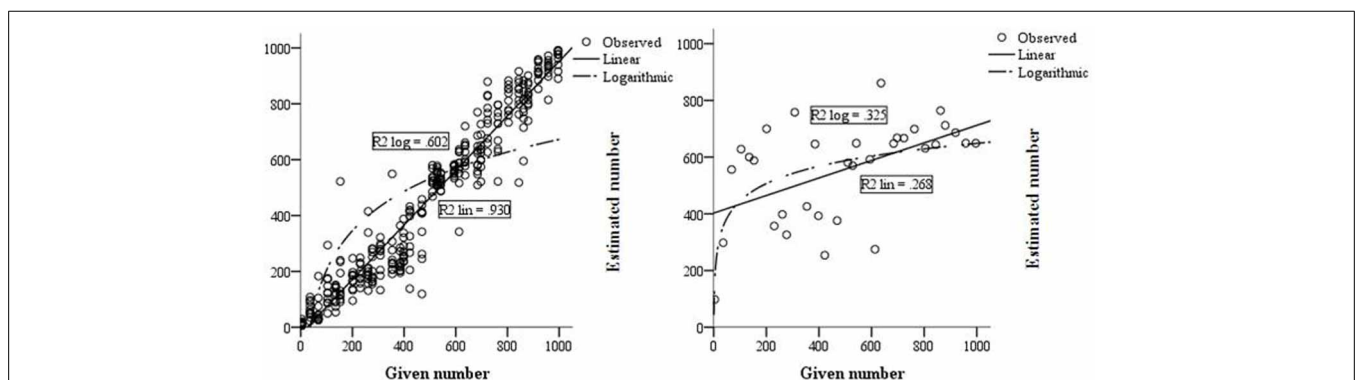
FIGURE 3 | Linear and logarithmic fit of L (right) and the control group (left) on the 0–1000 number line task.

However, in 40% of the cases no fixation pattern could be distinguished (i.e., 78, 18, 27, 83, 46, 14, 61, 74, 66, 32, 19, 80). L fixated at the middle, tens, and units separately in several differing sequences. The percentages for the control group are displayed in **Table 3**, illustrating the differences in strategy use. L used an atypical strategy about as often as the control group, but her strategy use was significantly more often undefined than the control group average (**Table 3**). When looking at the type of two-digit numbers, it is not possible to identify a pattern of which numbers are processed in what way, or differences between L and the control children in what strategies are used for the estimation of certain numbers.

On the symbolic 0–1000 number line task, L showed in 0% of the cases a holistic processing or decomposed sequential processing strategy. L used a decomposed parallel processing strategy in 20% of the cases (i.e., 684, 354, 385, 958, 996, 763) and an atypical strategy in 23.3% of the cases (i.e., 594, 613, 844, 862, 723, 919, 201), with eye fixations starting at the smallest entity and then progressing toward the largest entity. Moreover, in 56.7% of the cases, no specific fixation pattern could be identified (i.e., 422, 261, 528, 542, 398, 510, 804, 277, 469, 230, 104, 153, 697, 135, 880, 308, 636). The eye movements were alternating several times between the middle, hundreds, tens, and units, and while lugging the lever randomly across the number line, L constantly kept checking the separate digits. As a result, it was not possible to identify a (functional) strategy in two thirds of the cases. The percentages for the control group are displayed in **Table 3**. Again, L's processing of three-digit numbers was significantly more often unidentifiable than in the control group.

Estimation strategies

Eye movement patterns of the symbolic 0–100 and 0–1000 number line trials were analyzed in a qualitative way to discern information on L's strategy use in estimating numbers on a number line. Eye-tracking videos of the control children were also viewed and coded. Trials were coded as *counting-up* when a child used the beginning of the line as a reference point and counted up from there, and *counting-down* when the end of the line was used as a reference point. Trials in which the child used the midpoint

of the line and counted up or down from there were coded as *midpoint*. Finally, trails were coded as *undefined* when no clear estimation strategy could be determined from the eye movement patterns, for example because no reference point was used. If an estimation strategy was evaluated as inadequate, for example when a child used the endpoint strategy for number 18, that trial was double coded as *dysfunctional*. It was expected that L, having severe mathematical problems, would show more dysfunctional and undefined strategies than the control group, as she might make less use of reference points and gaze more at the reference point in the middle of the line (see Van't Noordende and Kolkman, 2013). On the symbolic 0–100 task, L used the beginning of the line as a reference point in 24.2% of the cases (i.e., 10, 3, 43, 9, 72, 32, 28, 5). For the numbers 43, 72, and 32, this strategy was coded as dysfunctional. She used the *midpoint strategy* in 30.3% of the trials (i.e., 53, 46, 14, 24, 74, 66, 87, 41, 57, 60) of which 14, 24, and 87 were estimated with a dysfunctional strategy. The numbers 78, 96, 91, 80, and 99 were all estimated in a functional manner with the *counting-down* strategy. L's eye movement patterns did not reflect an identifiable strategy in 30.3% of the cases (i.e., 18, 27, 64, 83, 49, 61, 89, 37, 34, 19). Instead, her eyes rapidly went back and forth across the number line without seemingly entailing a clear goal. L used a dysfunctional strategy in 26.1% of the identifiable trials. **Table 4** illustrates the differences in strategy use between L and the control group: it becomes clear that L's strategy use was significantly more often dysfunctional than the control group average. These results can be summarized in a graphical display, called a heat map. **Figure 4** summarizes all eye fixations of L (top) and three randomly chosen control children (bottom) during the symbolic 0–100 task in these heat maps. L's graph perfectly illustrates the randomness of her strategy use, whereas the maps of the control children consistently show the three distinguishable reference points (begin, middle, end).

For the symbolic 0–1000 task, L seldom employed the *counting-up* and *counting-down* strategy. However, as was expected, she utilized the middle reference point in two thirds of the trials (i.e., 261, 528, 542, 398, 594, 510, 277, 469, 684, 844, 104, 68, 385, 153, 723, 697, 958, 135, 996, 201, 636, 763), though this appeared to be a dysfunctional strategy in 50% of the cases

Table 3 | Overview of the processing strategies used (in percentages) by the child with dyscalculia and the control group (including range of percentages) in the symbolic 0–100 and 0–1000 number line tasks including *p*-values and effect sizes.

Strategy	Child with dyscalculia		Control group[a]		Test results			
	0–100	0–1000	0–100	0–1000	0–100		0–1000	
	%	%	%	%	*p*-value	z_{cc}	*p*-value	z_{cc}
Holistic	26.7	0.0	49.7 (30.0–73.3)[b]	22.0 (0.0–56.7)[b]	.057	−1.84	.144	−1.18
Decomposed parallel	13.3	20.0	21.3 (10.0–46.7)[b]	55.3 (20.0–76.7)[b]	.286	−0.62	.067	−1.72
Decomposed sequential	10.0	0.0	11.0 (6.7–23.3)[b]	1.7 (0.0–10.0)[b]	.461	−0.11	.314	−0.53
Other	10.0	23.3	12.0 (0.0–30.0)[b]	15.0 (3.0–30.0)[b]	.421	−0.22	.184	0.99
Undefined	40.0	56.7	6.0 (0.0–13.3)[b]	6.0 (0.0–26.7)[b]	.000	6.59	.000	5.90

[a] *n = 10, numbers represent the mean percentage of the 10 control children.*

[b] *Range of the percentages of the 10 control children, see appendix for individual results of the control children. Z_{cc}, obvious direct analog of Cohen's d, subscript representing "case-controls."*

Table 4 | Overview of the estimation strategies used (in percentages) by the child with dyscalculia and the control group (including range of percentages) in the symbolic 0–100 and 0–1000 number line tasks including *p*-values and effect sizes.

Strategy	Child with dyscalculia		Control group[a]		Test results			
	0–100	0–1000	0–100	0–1000	0–100		0–1000	
	%	%	%	%	*p*-value	z_{cc}	*p*-value	z_{cc}
Counting-up	24.2	3.0	22.4 (15.2–33.3)[b]	24.8 (12.1–36.4)[b]	.397	0.28	.016	−2.67
Midpoint	30.3	67.0	36.4 (21.2–42.4)[b]	31.5 (15.2–45.5)[b]	.193	−0.96	.002	3.87
Counting-down	15.2	9.1	25.1 (12.1–33.3)[b]	23.9 (12.1–39.4)[b]	.097	−1.48	.053	−1.88
Undefined	30.3	21.2	18.8 (6.1–42.3)[b]	19.1 (3.0–36.4)[b]	.193	0.59	.421	0.21
Dysfunctional	26.1	42.3	8.8 (0.0–19.4)[b]	11.1 (4.0–21.9)[b]	.010	2.97	.007	3.16

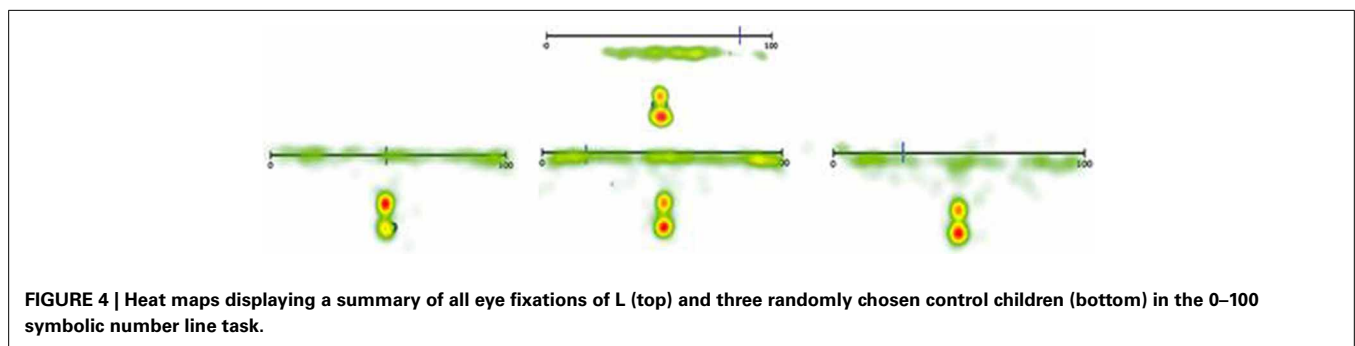[a]*n = 10, numbers represent the mean percentage of the 10 control children.*

[b]*Range of the percentages of the 10 control children, see appendix for individual results of the control children. Z_{cc}, obvious direct analog of Cohen's d, subscript representing "case-controls."*

FIGURE 4 | Heat maps displaying a summary of all eye fixations of L (top) and three randomly chosen control children (bottom) in the 0–100 symbolic number line task.

(e.g., 153, 996). Moreover, 21.2% of the trials were coded as *undefined*, as L made no use of any reference points. In total, she used a dysfunctional strategy in 42.3% of the cases, which was significantly more often than the control group (**Table 4**). This might illustrate a deficit in adapting strategies to a specific target number, making her number line estimates appear "random." Again, this seemingly random strategy employment can be illustrated by the heat maps summarizing all eye fixations on the symbolic 0–1000 task (**Figure 5**).

The results show that different number processing as well as estimation strategies can indeed be discerned using eye-tracking data. Moreover, it is possible to find differences in presence or absence of number processing strategies between children with typically developing numerical skills and children with a number processing deficit based on eye-tracking data. Although this study is exploratory in nature, these results provide a first indication of the possible value of eye-tracking data in combination with number line tasks to discriminate between typical and atypical numerical development in children in future diagnostic procedures.

DISCUSSION

In this study, the eye movements of a child (L) with DD on a series of number line tasks were compared to those of TD children, in order to determine whether a combination of number line tasks and eye-tracking data can be used to discriminate

between TD children and children with a number processing deficit. This case study has provided some important insights in the (non-)numerical processing strategies of children with DD in contrast to TD children. Performance characteristics of L quantitatively and qualitatively differed from the TD children, as there was a high discrepancy in accuracy and strategy use between L and the control group. This evidence suggests that eye-tracking data in combination with number line estimation tasks might be a promising tool in diagnosing dyscalculia in children.

The hypothesis stating that the eye-tracking data of L would indicate that her number representations were logarithmic rather than linear was partly confirmed by the data. In contrast to the 0–100 non-symbolic and 0–1000 symbolic task, her number representations on the 0–100 symbolic line seemed linear rather than logarithmic. The estimations of the control children were better explained by a linear fit on all tasks. L also displayed significantly lower accuracy in estimating numbers than the control group, meaning that her estimations were more distanced from the target number. These results are consistent with her diagnosis dyscalculia. In line with previous research, it might be concluded that the number representations of a child with dyscalculia around age 10 are less precise and therefore similar to younger children (Piazza et al., 2010). However, these results partly contradict findings of Van't Noordende and Kolkman (2013), who investigated number line estimations in children
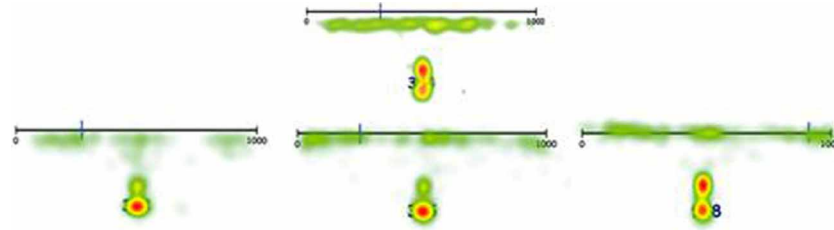
FIGURE 5 | Heat maps displaying a summary of all eye fixations of L (top) and three randomly chosen control children (bottom) in the 0–1000 symbolic number line task.

with math disabilities and concluded that number representations of both atypically and a TD children were better represented by a linear fit.

The second hypothesis, stating that L would show more atypical and more unidentifiable fixation patterns than the control group when processing two-digit and three-digit numbers, was largely confirmed. L showed 65% more atypical strategies than the control group on the 0–1000 number line task. In addition, L showed significantly more unidentifiable fixation patterns on the symbolic 0–100 task and nine times more unidentifiable fixation patterns on the 0–1000 task. These results provide important insights in the quality of magnitude representations in children with and without a specific number processing deficit. Eye-tracking data can be a useful tool to visualize the way a child processes multi-digit numbers and therefore creates the opportunity to take a step forward toward a diagnostic process based on cognitive rather than behavioral measures, or at least a combination of both. Because of the exploratory nature of this study, it is not possible to make any statements about which processing strategies indicate typical or atypical development. Future research is needed in order to explore whether TD children, for example, make more use of holistic processing than children with dyscalculia. It can, however, be stated that L made less use of identifiable strategies. Moreover, the relatively high percentages of trials that were coded as "other" in both L and the control group indicate that there might be another functional strategy that can be used to process multi-digit numbers. Future research could shed more light on the functionality and overlap or discrimination between different number processing strategies.

Finally, the third hypothesis stated that L would show more dysfunctional and more undefinable strategies than the control group when estimating numbers on a number line. This was also largely confirmed by the data. L exhibited a dysfunctional strategy significantly more often than the control group and also made less use of reference points in the symbolic 0–100 line (i.e., undefined strategy). Moreover, in the symbolic 0–1000 task, she gazed significantly more often (in 67% of the cases) at the midpoint of the line than the control group. These results are in line with previous research on children with math difficulties (Van't Noordende and Kolkman, 2013). The findings implicate that a child with dyscalculia might have more difficulty with flexibly adapting strategies to the target number than TD children, due to a deficit in the ability to connect a number symbol

to a non-symbolic magnitude (i.e., mapping deficit; Dehaene, 2001). The visuospatial memory deficit that was found in L might also partly explain her lower performance on the number line tasks used in this case study (Kolkman et al., under review).

Results of this study have to be interpreted in the light of some methodological caveats. For example, the number line tasks and eye-tracking methodology used in this study have not yet been validated. Hence, it might be that the number sense tasks or the eye-tracking data measure other aspects (i.e., measurement error) in addition to actual competence (Schneider et al., 2008). Moreover, in a small percentage of the total trials (0.5%) calibration errors occurred. This comes to pass when the eye movements are not monitored correctly by the computer and do not appear in the eye-tracking video. Consequently, these trials had to be coded as unidentifiable. Future research focusing on the validity and reliability of these measures (i.e., both eye-tracking methodology and number sense tasks) is essential.

Another caveat relates to the "n = 1" nature of this study: only one child with a numerical processing deficit was included. Accordingly, no firm conclusions can be drawn from these data. Future research with a larger empirical sample should investigate the eye fixation patterns in other children with dyscalculia, especially since several subtypes of math disabilities have been described (Wilson and Dehaene, 2007; Rubinsten and Henik, 2009). As a consequence, it might become possible to include eye-tracking measures and number line tasks as an additional tool in the diagnostic process of dyscalculia in the future. This would be of great value, since most current diagnostic instruments are only focused on the behavioral level.

Further theory development on number processing and estimation strategies is needed since a relatively high percentage of the trials in the present study was coded as "other" or "undefined." The nature of the differences in unidentifiable strategies between L and the control children is not clear. The existence of unidentifiable eye fixation patterns indicates that additional, yet to be defined, strategies might exist. For example, concerning the control group, absence of strategy use but quick estimation might indicate a certain level of automation: a child possesses a precise representation in long term memory of where some of the numbers, e.g., 50, belong on the number line and therefore does not need to employ a specific decoding strategy.

REFERENCES

Alloway, T. P. (2007). *Automated Working Memory Assessment (AWMA).* Amsterdam: Pearson.

Alloway, T. P., Gathercole, S. E., Kirkwood, H., and Elliot, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Dev.* 80, 606–621. doi: 10.1111/j.1467-8624.2009.01282.x

Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 41, 189–201. doi: 10.1037/0012-1649.41.6.189

Butterworth, B., Varma, S., and Laurillard, D. (2011). Dyscalculia: from brain to education. *Science* 332, 1049–1053. doi: 10.1126/science.1201536

Crawford, J. R., and Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia* 40, 1196–1208. doi: 10.1016/S0028-3932(01)00224-X

Crawford, J. R., Garthwaite, P. H., and Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: rationale, methods, implementations, and proposed reporting standards. *Cogn. Neuropsychol.* 27, 245–260. doi: 10.1080/02643294.2010.513967

Crawford, J. R., and Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *Clin. Neuropsychol.* 12, 482–486. doi: 10.1076/clin.12.4.482.7241

De Visscher, A., and Noel, M.-P. (2012). A case study of arithmetic facts dyscalculia caused by a hypersensitivity-to-inference memory. *Cortex* 49, 50–70. doi:10.1016/j.cortex.2012.01.003

De Weerdt, F., Desoete, A., and Roeyers, H. (2013). Working memory in children with reading disabilities and/or mathematical disabilities. *J. Learn. Disabil.* 46, 461–472. doi: 10.1177/0022219412455238

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics.* Oxford: Oxford University Press.

Dehaene, S. (2001). Précis of the number sense. *Mind Lang.* 16, 16–36. doi: 10.1111/1468-0017.00154

Dehaene, S., Dupoux, E., and Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 626–641. doi: 10.1037/0096-1523.16.3.626

Desoete, A., Ceulemans, A., De Weerdt, F., and Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? Findings from a longitudinal study. *Br. J. Educ. Psychol.* 82, 64–81. doi: 10.1348/2044-8279.002002

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., and Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year old children: evidence for a segmented linear model. *J. Exp. Child Psychol.* 99, 1–17. doi: 10.1016/j.jecp.2007.08.006

Geary, D. C., Hoard, M. K., Nugent, L., and Bailey, D. H. (2012). Mathematical cognitive deficits in children with learning disabilities and persistent low achievement: a five-year prospective study. *J. Educ. Psychol.* 104, 206–233. doi: 10.1037/a0025398

Heine, A., Thaler, V., Tamm, S., Hawelka, S., Schneider, M., Torbeyns, J., et al. (2010). What the eyes already 'know': using eye movement measurement to tap into children's implicit numerical magnitude representations. *Infant Child Dev.* 19, 175–186. doi: 10.1002/icd.640

Janssen, J., Scheltens, F., and Kraemer, J. M. (2005). *Leerling- en onderwijsvolgsysteem rekenen-wiskunde [Student and Education Tracking System Arithmetic-Mathematics].* Arnhem: Cito.

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

Korvorst, M., and Damian, M. F. (2008). The differential influence of decades and units on multi-digit number comparison. *Q. J. Exp. Psychol.* 61, 1250–1264. doi: 10.1080/17470210701503286

Kroesbergen, E. H., Van de Rijt, B. A. M., and Van Luit, J. E. H. (2007). "Working memory and early mathematics: possibilities for early identification of mathematics learning disabilities," in *Advances in Learning and Behavioral Disabilities.* Vol. 20: *International Perspectives,* eds T. E. Scruggs and M. A. Mastropieri (Oxford: Elsevier), 1–19.

Kucian, K., Loenneker, T., Dietrich, T., Dosch, M., Martin, E., and Von Aster, M. (2006). Impaired neural networks for approximate calculation in dyscalculic children: a functional MRI study. *Behav. Brain Funct.* 2, 1–17. doi: 10.1186/1744-9081-2-31

Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011). Impaired acquity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Dev.* 82, 1224–1237. doi: 10.1111/j.1467-8624.2011.01608.x

Moeller, K., Fischer, M. H., Nuerk, H. C., and Willmes, K. (2009a). Sequential or parallel decomposed processing of two-digit numbers? Evidence from eye tracking. *Q. J. Exp. Psychol.* 62, 323–334. doi: 10.1080/17470210801946740

Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2009b). Children's early mental number line: logarithmic or decomposed linear? *J. Exp. Child Psychol.* 103, 503–515. doi: 10.1016/j.jecp.2009.02.006

Newman, R. S., and Berger, C. F. (1984). Children's numerical estimation: flexibility in the use of counting. *J. Educ. Psychol.* 76, 55–64. doi: 10.1037/0022-0663.76.1.55

Nuerk, H. C., Moeller, K., Klein, E., Willmes, K., and Fischer, M. H. (2011). Extending the mental number line: a review of multi-digit number processing. *J. Psychol.* 219, 3–22. doi: 10.1027/2151-2604/a000041

Nuerk, H.-C., and Willmes, K. (2005). On the magnitude representations of two-digit numbers. *Psychol. Sci.* 47, 52–72.

Opfer, J. E., and Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cogn. Psychol.* 55, 169–195. doi:10.1016/j.cogpsych.2006.09.002

Passolunghi, M. C., and Mammarella, I. C. (2012). Selective spatial working memory impairment in children with arithmetic learning disabilities. *J. Learn. Disabil.* 45 341–350. doi: 10.1177/0022219411400746

Petitto, A. L. (1990). Development of numberline and measurement concepts. *Cogn. Instr.* 7, 55–78. doi: 10.1207/s1532690xci0701_3

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012

Poltrock, S. E., and Schwartz, D. R. (1984). Comparative judgement of multi-digit numbers. *J. Exp. Psycho. Learn. Mem. Cogn.* 10, 32–45. doi: 10.1037/0278-7393.10.1.32

Rotzer, S., Kucian, K., Martin, E., Von Aster, M., Klaver, P., and Loenneker, T. (2008). Optimized voxel-based morphometry in children with developmental dyscalculia. *Neuroimage* 39, 417–422. doi: 10.1016/j.neuroimage.2007.08.045

Rotzer, S., Loenneker, T., Kucian, K., Martin, E., Klaver, P., and Von Aster, M. (2009). Dysfunctional neural network of spatial working memory contributes to developmental dyscalculia. *Neuropsychologia* 47, 2859–2865. doi: 10.1016/j.neuropsychologia.2009.06.009

Rubinsten, O., and Henik, A. (2009). Developmental dyscalculia: heterogeneity might not mean different mechanisms. *Trends Cogn. Sci.* 13, 92–99. doi: 10.1016/j.tics.2008.11.002

Schneider, M., Heine, A., Thaler, V., Torbeyns, J., De Smedt, B., Verschaffel, L., et al. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cogn. Dev.* 23, 424–437. doi: 10.1016/j.cogdev.2008.07.002

Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi:10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., Thompson, C., and Opfer, J. E. (2009). The logarithmic-to-linear shift: one learning sequence, many tasks, many time scales. *Mind Brain Educ.* 3, 143–150. doi:10.1111/j.1751-228X.2009.01064.x

Sullivan, J. L., Juhasz, B. J., Slattery, T. J., and Barth, H. C. (2011). Adults' number-line estimation strategies: evidence from eye movements. *Psychon. Bull. Rev.* 18, 557–563. doi:10.3758/s13423-011-0081-1

Toll, S. W. M., Van der Ven, S. H. G., Kroesbergen, E. H., and Van Luit, J. E. H. (2011). Executive functions as predictors of math learning disabilities. *J. Learn. Disabil.* 44, 521–532. doi: 10.1177/0022219410387302

Van Luit, J. E. H., Bloemert, J., Ganzinga, E. G., and Mönch, M. E. (2012). *Protocol*

Dyscalculie: Diagnostiek voor Gedragsdeskundigen [*Protocol Dyscalculia: Diagnostics for Psychologists*]. Doetinchem: Graviant.

Van't Noordende, J. E., and Kolkman, M. E. (2013). Getallenlijnschatten door kinderen met en zonder rekenproblemen: accuratesse, representaties en strategiegebruik [Number line estimation in children with and without math learning problems: accuracy, representations, and strategy use]. *Orthopedagogiek:* *Onderzoek en Praktijk* 52, 322–335.

Wilson, A. J., and Dehaene, S. (2007). "Number sense and developmental dyscalculia," in *Human Behavior, Learning and the Developing Brain: Atypical Development*, eds D. Coch, G. Dawson, and K. Fischer (New York, NY: Guilford Publications), 212–238.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

Table A1 | Overview of the individual working memory scores of the child with dyscalculia and the control children.

	Dot matrix			Non-word recall		
	Raw score	Standard score	Percentile	Raw score	Standard score	Percentile
L	–	63	1	–	89	22
CONTROL CHILD						
1	21	104	58	16	117	85
2	21	94	32	15	106	64
3	28	126	96	12	93	29
4	24	104	55	11	88	19
5	23	103	53	15	106	64
6	30	135	99	13	97	39
7	18	81	11	13	97	39
8	29	130	98	13	97	39
9	32	144	100	15	106	64
10	24	108	67	13	97	39

The standard scores have a mean of 100 and a standard deviation of 15.

Table A2 | Overview of the individual logarithmic and linear fit measures of the child with dyscalculia and the control children on the non-symbolic 0–100 and symbolic 0–100 and 0–1000 number line tasks.

	NS 0–100		S 0–100		S 0–1000	
	Linear	Logarithmic	Linear	Logarithmic	Linear	Logarithmic
L	0.688	0.705	0.847	0.698	0.268	0.325
CONTROL CHILD						
1	0.868	0.799	0.848	0.752	0.854	0.600
2	0.748	0.848	0.973	0.796	0.967	0.626
3	0.740	0.634	0.960	0.764	0.947	0.584
4	0.794	0.619	0.975	0.784	0.936	0.592
5	0.696	0.629	0.964	0.773	0.944	0.585
6	0.892	0.816	0.980	0.800	0.934	0.557
7	0.818	0.691	0.975	0.791	0.967	0.629
8	0.783	0.727	0.976	0.787	0.964	0.640
9	0.935	0.776	0.972	0.815	0.887	0.607
10	0.922	0.765	0.963	0.775	0.943	0.620

Table A3 | Overview of the individual mean percent absolute errors of the child with dyscalculia and the control children on the non-symbolic 0–100 and symbolic 0–100 and 0–1000 number line tasks.

Number line	Absolute error (%)		
	NS 0–100	S 0–100	S 0–1000
L	17.7	10.7	20.1
CONTROL CHILD			
1	10.5	7.0	8.0
2	19.3	3.5	4.2
3	16.9	5.0	6.4
4	11.4	3.9	6.1
5	12.0	5.7	6.4
6	13.8	3.6	6.0
7	15.4	4.5	5.8
8	16.0	3.6	4.9
9	6.3	4.0	7.0
10	8.4	4.5	6.1

Table A4 | Overview of the individual percentages of the processing strategies in the symbolic 0–100 number line task of the child with dyscalculia and the control children.

Strategy	Holistic (%)	Decomposed parallel (%)	Decomposed sequential (%)	Other (%)	Undefined (%)
L	26.7	13.3	10.0	10.0	40.0
CONTROL CHILD					
1	40.0	20.0	6.7	30.0	3.3
2	43.3	10.0	20.0	13.3	13.3
3	50.0	20.0	23.3	6.7	0.0
4	53.5	23.3	6.7	6.7	10.0
5	46.7	10.0	16.7	13.3	13.3
6	73.3	23.3	0.0	3.3	0.0
7	60.0	10.0	13.3	16.7	0.0
8	60.0	10.0	23.3	0.0	6.7
9	30.0	40.0	0.0	23.3	6.7
10	40.0	46.7	0.0	6.7	6.7

Table A5 | Overview of the individual percentages of the processing strategies in the symbolic 0–1000 number line task of the child with dyscalculia and the control children.

Strategy	Holistic (%)	Decomposed parallel (%)	Decomposed sequential (%)	Other (%)	Undefined (%)
L	0.0	20.0	0.0	23.3	56.7
CONTROL CHILD					
1	33.3	56.7	0.0	10.0	0.0
2	36.7	40.0	3.3	16.7	3.3
3	0.0	70.0	0.0	3.3	26.7
4	56.7	20.0	0.0	13.3	10.0
5	6.7	73.3	0.0	6.7	13.3
6	13.3	73.3	0.0	13.3	0.0
7	23.3	66.7	0.0	10.0	0.0
8	36.7	26.7	10.0	23.3	3.3
9	13.3	50.0	3.3	30.0	3.3
10	0.0	76.7	0.0	23.3	0.0

Table A6 | Overview of the individual percentages of the estimation strategies in the symbolic 0–100 number line task of the child with dyscalculia and the control children.

Strategy	Counting-up (%)	Midpoint (%)	Counting-down (%)	Undefined (%)	Dysfunctional (%)
L	24.2	30.3	15.2	30.3	26.1
CONTROL CHILD					
1	18.2	30.3	24.2	27.3	8.3
2	30.3	30.3	33.3	6.1	19.4
3	18.2	36.4	33.3	12.1	13.8
4	27.3	27.3	27.3	15.2	0.0
5	15.2	39.4	12.1	33.3	4.5
6	18.2	42.4	21.2	18.2	14.8
7	18.2	21.2	18.2	42.3	10.5
8	27.3	39.4	24.2	9.1	10.0
9	33.3	33.3	27.3	6.1	3.2
10	18.2	33.3	30.3	18.2	7.4

Table A7 | Overview of the individual percentages of the estimation strategies in the symbolic 0–1000 number line task of the child with dyscalculia and the control children.

Strategy	Counting up (%)	Midpoint (%)	Counting-down (%)	Undefined (%)	Dysfunctional (%)
L	3.0	67.0	9.1	21.2	42.3
CONTROL CHILD					
1	12.1	30.3	27.3	30.3	13.0
2	30.3	30.3	27.2	12.1	6.9
3	21.1	30.3	12.1	36.4	14.3
4	24.2	30.3	21.2	18.2	4.0
5	33.3	15.2	39.4	12.1	4.8
6	33.3	21.2	24.2	21.2	15.4
7	21.2	39.4	18.2	21.2	11.5
8	21.2	42.4	24.2	12.1	10.3
9	36.4	45.5	15.2	3.0	21.9
10	15.2	30.3	30.3	24.2	14.3

Dyscalculia from a developmental and differential perspective

Liane Kaufmann[1]*, Michèle M. Mazzocco[2], Ann Dowker[3], Michael von Aster[4,5,6], Silke M. Göbel[7], Roland H. Grabner[8], Avishai Henik[9], Nancy C. Jordan[10], Annette D. Karmiloff-Smith[11], Karin Kucian[6], Orly Rubinsten[12], Denes Szucs[13], Ruth Shalev[14] and Hans-Christoph Nuerk[15,16,17]

[1] Department of Psychiatry and Psychotherapy A, General Hospital, Hall in Tyrol, Austria
[2] Institute of Child Development, University of Minnesota, Minneapolis, MN, USA
[3] Department of Experimental Psychology, University of Oxford, Oxford, UK
[4] Department for Child and Adolescent Psychiatry, German Red Cross Hospitals, Berlin, Germany
[5] Department for Clinical Psychology, University of Potsdam, Potsdam, Germany
[6] MR-Center, University Children's Hospital, Zürich, Switzerland
[7] Department of Psychology, University of York, York, UK
[8] Department of Psychology, Georg-August-University of Goettingen, Goettingen, Germany
[9] Department of Psychology, Faculty of Humanities and Social Sciences, Ben Gurion University of the Negev, Negev, Israel
[10] School of Education, University of Delaware, Newark, DE, USA
[11] Centre for Brain and Cognitive Development, Department of Psychological Sciences, Birkbeck, University of London, London, UK
[12] Edmond J. Safra Brain Research Center for the Study of Learning Disabilities, Department of Learning Disabilities, University of Haifa, Haifa, Israel
[13] Department of Psychology, Centre for Neuroscience in Education, University of Cambridge, Great Britain, UK
[14] Neuropediatric Unit, Shaare Zedek Medical Center, Jerusalem, Israel
[15] Department of Psychology, University of Tuebingen, Tuebingen, Germany
[16] IWM-KMRC, Knowledge Media Research Center, Tuebingen, Germany
[17] LEAD Graduate School, Tuebingen, Germany
*Correspondence: liane.kaufmann@tilak.at

Developmental dyscalculia (DD) and its treatment are receiving increasing research attention. A *PsychInfo* search for peer-reviewed articles with *dyscalculia* as a title word reveals 31 papers published from 1991–2001, versus 74 papers published from 2002–2012. Still, these small counts reflect the paucity of research on DD compared to dyslexia, despite the prevalence of mathematical difficulties. In the UK, 22% of adults have mathematical difficulties sufficient to impose severe practical and occupational restrictions (Bynner and Parsons, 1997; National Center for Education Statistics, 2011). It is unlikely that all of these individuals with mathematical difficulties have DD, but criteria for defining and diagnosing dyscalculia remain ambiguous (Mazzocco and Myers, 2003). What is treated as DD in one study may be conceptualized as another form of mathematical impairment in another study. Furthermore, DD is frequently—but, we believe, mistakenly- considered a largely homogeneous disorder. Here we advocate a differential and developmental perspective on DD focused on identifying behavioral, cognitive, and neural sources of individual differences that contribute to

our understanding of what DD *is* and what it is *not*.

HETEROGENEITY IS A FEATURE OF DD

DD is not synonymous with all forms of arithmetic and mathematical difficulties[1]. Here we emphasize that DD is characterized by severe arithmetic difficulties and accounts for only a subset of individuals with arithmetic difficulties [see Figure 2 in Kaufmann and von Aster (2012)]. In studies including children with various manifestations of arithmetic difficulties, true deficits of DD are likely to be masked because DD represents only a minority of children in these samples (Murphy et al., 2007; LeFevre et al., 2010). Any theory of DD must account for differences between DD and individual differences in arithmetic in the general population. Kaufmann and Nuerk (2005) claimed that, "... average arithmetic development does not pursue

---

[1]The terms "arithmetic" and "mathematical" are not synonymous as the former refers to computational skills (i.e., processing of basic arithmetical operations such as addition/subtraction/multiplication) and the latter encompasses other aspects of numerical thinking such as algebra, geometry, etc.

a straight, fully predictable course of acquisition, but rather can be characterized by quite impressive individual differences" (Siegler, 1995; Dowker, 2005). Arithmetic ability consists of many components [e.g., memorizing facts, executing procedures, understanding, and using arithmetical principles (Desoete et al., 2004; Dowker, 2005, 2008)], each subject to individual differences that continue into adulthood (Dowker, 2005; Kaufmann et al., 2011a) and may contribute to the reported prevalence of low numeracy (Geary et al., 2013). These individual differences must be considered when defining DD, because assumptions about a single core deficit (e.g., Butterworth, 2005) do not support the range of clinical manifestations of DD.

Moreover, heterogeneity of DD and other mathematics difficulties is also fostered by environmental factors, ranging from cultural factors (e.g., nature and extent of schooling, characteristics of the counting system) to the effects of pre-/postnatal illness or socio-emotional adversity (e.g., math anxiety). Hence, arithmetic difficulties may be associated with other learning disorders (i.e., dyslexia) or
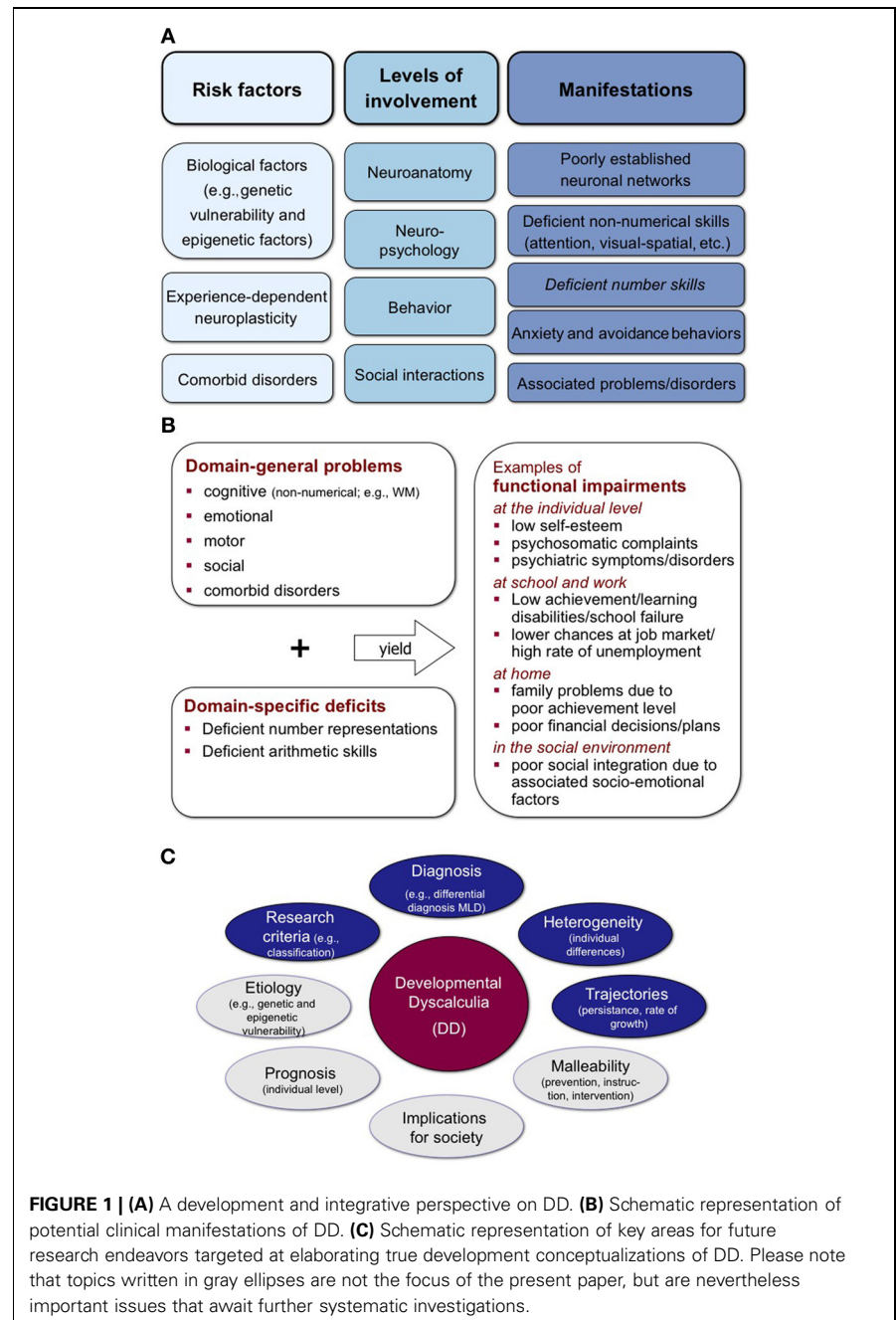
with various neuropsychiatric and pediatric disorders (e.g., attention-deficit hyperactivity-disorder/ADHD, epilepsy; Shalev and Gross-Tsur, 1993; Marzocchi et al., 2002; Kaufmann and Nuerk, 2008). Disentangling these types of arithmetic difficulties may be important given recent evidence that treating an underlying medical condition (i.e., attention disorder) may alleviate the arithmetic difficulties (Rubinsten et al., 2008).

Below, we emphasize the need for a developmental view on DD and suggest definitional criteria acknowledging its developmental nature, heterogeneous manifestations and distinctness from other forms of arithmetic/mathematical difficulties.

TOWARDS A DEVELOPMENTAL PERSPECTIVE ON DD

A developmental perspective enables us to trace pathways of parallel and/or sequential mechanisms at varying processing levels (neuroanatomical, neuropsychological, behavioral, interactional; **Figure 1A**). Important questions facing researchers include whether DD represents the extreme end of a continuum (or several continua) of mathematical ability or whether the arithmetic difficulties associated with DD are qualitatively different from more common mathematics difficulties. There is evidence to support each of these positions.

Arithmetic difficulties can reflect individual differences in both numerical and non-numerical functions. The numerical functions comprise many aspects of "number sense" such as spontaneous focusing on number (Hannula et al., 2010), comparing numerical quantities represented non-symbolically (e.g., as dot arrays; Piazza et al., 2010; Halberda et al., 2012), processing numbers symbolically (e.g., in Arabic notation; Stock et al., 2010), or linking non-symbolic representations to symbols such as number words and Arabic numerals (Rubinsten et al., 2002; Rubinsten and Henik, 2005; Bugden and Ansari, 2011). These individual differences in "number sense" may reflect variation in neural pathways involved in even quite rudimentary aspects of numerical cognition (e.g., single digit arithmetic: Price et al., 2013). Studies of functional activation during magnitude comparison reflect developmental variations over time (for respective

FIGURE 1 | (A) A development and integrative perspective on DD. **(B)** Schematic representation of potential clinical manifestations of DD. **(C)** Schematic representation of key areas for future research endeavors targeted at elaborating true development conceptualizations of DD. Please note that topics written in gray ellipses are not the focus of the present paper, but are nevertheless important issues that await further systematic investigations.

meta-analyses, see Houdé et al., 2010; Kaufmann et al., 2011b) and suggest variation in development *per se* rather than in comparable but delayed trajectories (Vogel and Ansari, 2012; Price et al., 2013).

Recently, Moeller et al. (2012) distinguished the following approaches: (i) DD is related to a numerical core deficit, (ii) DD subtypes exist due to domain-general processes, and (iii) DD subtypes exist due to domain-specific numerical deficits beyond the aforementioned core numerical deficit. The

core deficit hypothesis assumes that DD is a coherent syndrome mainly linked to neurofunctional peculiarities of the intraparietal sulcus (Butterworth, 2005). However, the heterogeneous clinical picture of DD (**Figure 1B**) is at odds with a single core deficit assumption (Mazzocco, 2007; Rubinsten and Henik, 2009). The second approach suggests that different subtypes can be distinguished on the basis of associated *domain-general deficits*. For instance, deficits in verbal (working) memory, semantic memory or

visual-spatial skills (Rourke and Conway, 1997; von Aster, 2000; Geary, 2004) and even in belief-laden logical reasoning (Morsanyi et al., 2013) reportedly influence arithmetic difficulties (although some results contradict any view of simple relationships between verbal/spatial discrepancies and arithmetical components; Dowker, 1998). Respective developmental calculation models acknowledging non-numerical influences have been proposed previously (von Aster and Shalev, 2007; Kaufmann et al., 2011b). Such domain-general cognitive deficits may account for individual differences in the clinical picture despite comparable core numerical deficits. Finally, *domain-specific numerical deficits* (Wilson and Dehaene, 2007) may reflect multiple and distinct genuinely numerical deficits specifically affecting magnitude representation, verbal number representations, arithmetic fact knowledge, visual-spatial number forms, ordinality, base-10-system, or finger representations of numbers (Temple, 1991; Mazzocco et al., 2011; Moeller et al., 2012).

CURRENT CHALLENGES RELATED TO DD CLASSIFICATION, DIAGNOSIS, AND RESEARCH CRITERIA

These aforementioned theoretical assumptions have important consequences for DD diagnosis and research. If, for instance, some children have severe problems in arithmetic fact retrieval but perform adequately on other numerical and arithmetic assessment tasks, they might not be classified as dyscalculic or even arithmetically impaired when assessments rely on a composite score comprising different numerical and arithmetic tasks. Deficits in one or few subsets that do not qualify for a DD diagnosis may still constitute severe problems for those children. In research designs, such delineated deficits might be undetected by group studies because averaging across participants and processes may mask deficits displayed by minorities (Siegler, 1987). The opposite risk also exists: children may be labeled, by themselves or others, as weak at arithmetic based on a specified difficulty despite average or high ability in other areas of arithmetic. This may lead to self-fulfilling prophecies or contribute to significant mathematics anxiety. Indeed, among young children, most studies suggest relatively

little relationship between anxiety and performance, while in older children and adults, the relationship is strong and bidirectional; anxiety affects performance, and poor performance leads to anxiety (e.g., Ashcraft and Kirk, 2001; Mazzone et al., 2007; Pixner and Kaufmann, 2013).

Another major challenge of research on DD is the extensive range seen in diagnostic criteria and assessment tools used, which may influence research results (Murphy et al., 2007; Moser Opitz and Ramseier, 2012; Devine et al., 2013). As discussed by Moeller et al. (2012), there is little agreement about which children belong in the target group (DD, mathematical learning disability, etc.). Methodological approaches vary in terms of the cut-off points for classification criteria (ranging from <10 to <35 percentiles), whether reported percentiles reflect standardized or sample-based rankings, or deviations based on the population means and SDs. When different approaches are used across studies, very different children are included in study samples, and thus different background characteristics may be controlled for. Even children with general cognitive deficits may be included if a significant discrepancy between average intellectual abilities and sub-average math skills is not required as definitional criterion (as requested by the current Diagnostic and Statistical Manual of Mental Disorders (DSM) (Ehlert et al., 2012).

A final major challenge concerns the actual differential diagnostic classification tasks used in studies examining DD. While some studies employ discrete numerical tasks (e.g., dot enumeration), other studies use standardized math tests that may involve logical reasoning or text comprehension. Hence, apparently contradictory results as to whether DD involves deficits in basic or more complex numerical abilities may stem from the use of different classification tasks across studies. Discrepant findings may also reflect different samples of children who are nevertheless all presumed to have DD. The need is for research on DD to be both comprehensive and comparable across studies, which calls for a consortium-based proposal to adopt international standard diagnostic tools that are comparable across countries, curricula and therefore studies, in addition to study-specific assessments (as applicable).

HOW DEVELOPMENTAL CONCEPTUALIZATIONS OF DD MAY GUIDE EDUCATIONAL AND THERAPEUTIC APPROACHES

Beyond its scientific value, developmental conceptualizations of DD are crucial in guiding effective educational and therapeutic strategies. Researchers must consider the utility and meaningfulness of their contributions to the public perception of DD (including perceptions of teachers and parents). For instance, neurodevelopmental disorders like DD are at least partially attributable to inherited genetic differences (Shalev et al., 2001; Kovas et al., 2007). Hence, when conceptualized as a homogeneous and inborn disorder, DD may be misinterpreted as immune to the effects of behavioral interventions. A developmental approach considers multiple factors interacting to contribute to manifestations of DD. Such an approach is adopted in the forthcoming DSM-V, which replaces the categorical DSM-IV definition of distinct learning disorders (reading/written expression/mathematics) with an overarching multi-dimensional diagnosis of "Specific Learning Disorders" that acknowledges distinct manifestations of learning difficulties in various academic domains. However, in the theoretical debate about domain-specific versus domain-general underpinnings of DD, it is important to recall that domain-general deficits early on in development may result in seemingly domain-specific deficits in later development, because the earlier deficits may be more relevant to the computational demands of one domain (e.g., number) while still affecting other domains albeit to a more subtle degree. The reverse may also be true: numerical deficits may manifest as domain general deficits in, for instance, attention or working memory when diagnostic tools draw on numerical stimuli.

While advocating a developmental and differential perspective on DD, we must also caution against over-relying on adult neuropsychological patients with acquired mathematics disorders as models of DD (Kaufmann and Nuerk, 2005; Ansari, 2010; Karmiloff-Smith et al., 2012). As Karmiloff-Smith (1998) explains, important differences exist between deficits that arise during development versus those resulting from damage to an

existing system. Therefore, we argue that (i) DD is a heterogeneous disorder resulting from individual differences in development or function at neuroanatomical, neuropsychological, behavioral, and interactional levels (**Figure 1A**), and that (ii) an understanding of these differences can facilitate DD diagnosis and intervention. The acknowledgement of individual differences characterizing DD calls for adequate methodological and differential diagnostic approaches, and adequate attention to the developmental component of DD (reflecting systematic inter- and intra-individual variations between age and skill levels) (**Figure 1C**). Solid developmental conceptualizations of DD may foster the acceptance of DD as a disorder and raise public awareness for the need to provide targeted educational, therapeutic, and structural support tailored to affected individuals (**Figure 1B**), as well as differentiating DD from other sources of difficulty in children underperforming in mathematics.

As a synopsis of our arguments, we propose the following preliminary definition of DD:

*Primary DD is a heterogeneous disorder resulting from individual deficits in numerical or arithmetic functioning at behavioral, cognitive/neuropsychological and neuronal levels. The term **secondary DD** should be used if numerical/arithmetic dysfunctions are entirely caused by non-numerical impairments (e.g., attention disorders)*[2].

Further, we postulate the following recommendations for primary DD (and its diagnosis):

- There is convincing evidence that basic numerical skills are impaired in DD. Therefore, purely educational (curricular) tests are not adequate to tap the characteristic numerical deficits associated with DD.
- DD is a heterogeneous disorder (like other neurodevelopmental disorders). Multi-dimensional assessments tracking different numerical representations and arithmetic processes should be used to evaluate response accuracy, speed, and strategies.
- Specific deficits in numerical subdomains are possible, even when overall dyscalculia test scores are unremarkable.

---

[2]Likewise, Geary (2007) distinguished primary and secondary biological routes to math learning disabilities.

- Arithmetic performance of children diagnosed with DD can be unstable over development and time; thus children who are reasonably close to formal DD criteria (usually scoring <10th percentile) should be retested within the following school semester/year. Conservatively, retesting is recommended if performance is <25th percentile.
- Currently, there is no evidence that focusing on discrepancies between numerical and general cognitive skills improves diagnostic accuracy or interventional outcomes.
- DD can be comorbid with other neurodevelopmental, psychiatric, and neuropediatric disorders that may affect the regulation of motor/executive/affective/socio-behavioral functioning and have to be considered for differential diagnosis.
- Educational and socio-emotional characteristics should be considered in diagnosing and ruling out DD.

REFERENCES

Ansari, D. (2010). Neurocognitive approaches to developmental disorders of numerical and mathematical cognition: the perils of neglecting development. *Learn. Indiv. Diff.* 20, 123–129. doi: 10.1016/j.lindif.2009.06.001

Ashcraft, M. H., and Kirk, E. P. (2001). The relationship among working memory, math anxiety and performance. *J. Exp. Psychol.* 130, 224–237. doi: 10.1037/0096-3445.130.2.224

Bugden, S., and Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition* 118, 32–44. doi: 10.1016/j.cognition.2010. 09.005

Butterworth, B. (2005). The development of arithmetical abilities. *J. Child Psychol. Psychiatry* 46, 3–18. doi: 10.1111/j.1469-7610.2004.00374.x

Bynner, J., and Parsons, S. (1997). *Does Numeracy Matter. Evidence from the National Child*

Development Study on the Impact of Poor Numeracy on Adult Life. London: Basic Skills Agency.

Desoete, A., Roeyers, H., and De Clercq, A. (2004). Children with mathematics learning disabilities in Belgium. *J. Learn. Disabil.* 37, 50–61. doi: 10.1177/00222194040370010601

Devine, A., Soltész, F., Nobes, A., Goswami, U., and Szucs, D. (2013). Gender differences in developmental dyscalculia depend on diagnostic criteria. *Learn. Instruc.* 27, 31–39. doi: 10.1016/j.learninstruc.2013.02.004

Dowker, A. (2005). *Individual Differences in Arithmetical Abilities: Implications for Psychology, Neuroscience and Education.* New York, NY: Psychology Press. doi: 10.4324/9780203324899

Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Dev. Sci.* 11, 650–654. doi: 10.1111/j.1467-7687.2008.00713.x

Dowker, A. D. (1998). "Individual differences in normal arithmetical development," in *The Development of Mathematical Skills,* ed C. Donlan (East Sussex, UK: Lawrence Erlbaum Associates), 275–302.

Ehlert, A., Schroeders, U., and Fritz-Stratmann, A. (2012). Kritik am Diskrepanzkriterium in der Diagnostik von Legasthenie und Dyskalkulie. [Criticism of the discrepancy criterion in the diagnosis of dyslexia and dyscalculia]. *Lern. Lernstör.* 1, 169–184. (Full article in German, extended abstract in English). doi: 10.1024/2235-0977/a000018

Geary, D. C. (2004). Mathematics and learning disabilities. *J. Learn. Disabil.* 37, 4–15. doi: 10.1177/00222194040370010201

Geary, D. C. (2007). "Educating the evolved mind: conceptual foundations for an evolutionary educational psychology," in *Educating the Evolved Mind*, Vol. 2, eds J. S. Carlson and J. R. Levin (*Psychological perspectives on contemporary educational issues).* Greenwich, CT: Information Age), 1–99.

Geary, D. C., Hoard, M. K., Nugent, L., and Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PLoS ONE* 8:e54651. doi: 10.1371/journal.pone.0054651

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., and Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11116–11120. doi: 10.1073/pnas.1200196109

Hannula, M. M., Lepola, J., and Lehtinen, E. (2010). Spontaneous focusing on numerosity as a domain-specific predictor of arithmetical skills. *J. Exp. Child Psychol.* 107, 394–406. doi: 10.1016/j.jecp.2010.06.004

Houdé, O., Rossi, S., Lubin, A., and Joliot, M. (2010). Mapping numerical processing, reading, and executive functions in the developing brain: an fMRI meta-analysis of 52 studies including 842 children. *Dev. Sci.* 13, 876–885. doi: 10.1111/j.1467-7687.2009.00938.x

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends Cogn. Sci.* 2, 389–398. doi: 10.1016/S1364-6613(98)01230-3

Karmiloff-Smith, A., D'Souza, D., Dekker, T. M., Van Herwegen, J., Xu, F., Rodic, M., et al. (2012). Genetic and environmental vulnerabilities in children with neurodevelopmental

disorders. *Proc. Natl. Acad. Sci. U.S.A.* 109, 17261–17265.

Kaufmann, L., and Nuerk, H.-C. (2005). Numerical development: current issues and future perspectives. *Psychol. Sci.* 47, 142–170.

Kaufmann, L., and Nuerk, H.-C. (2008). Basic number processing deficits in ADHD: a broad examination of elementary and complex number processing skills in 9 to 12 year-old children with ADHD-C. *Dev. Sci.* 11, 692–699. doi: 10.1111/j.1467-7687.2008.00718.x

Kaufmann, L., Pixner, S., and Göbel, S. (2011a). Finger usage and arithmetic in adults with math difficulty: evidence from a case report. *Front. Cogn.* 2:254. doi: 10.3389/fpsyg.2011.00254

Kaufmann, L., Wood, G., Rubinsten, O., and Henik, A. (2011b). Meta-analysis of developmental fMRI studies investigating typical and atypical trajectories of number processing and calculation. *Dev. Neuropsychol.* 36, 763–787. doi: 10.1080/87565641.2010.549884

Kaufmann, L., and von Aster, M. (2012). The diagnosis and management of dyscalculia. *Dtsch. Arztebl. Int.* 109, 767–778.

Kovas, Y., Haworth, C. M. A., Petrill, S. A., Dale, P. S., and Plomin, R. (2007). Overlap and specificity of genetic and environmental influences on mathematics and reading disabilitiesin 10-year-old twins. *J. Child Psychol. Psychiatry* 48, 914–922. doi: 10.1111/j.1469-7610.2007.01748..x

LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: longitudinal predictors of performance. *Child Dev.* 81, 1753–1767. doi: 10.1111/j.1467-8624.2010.01508.x

Marzocchi, G. M., Lucangeli, D., De Meo, T., Fini, F., and Cornoldi, C. (2002). The disturbing effect of irrelevant information on arithmetic problem solving in inattentive children. *Dev. Neuropsychol.* 21, 73–92. doi: 10.1207/S15326942DN2101_4

Mazzocco, M. M. M. (2007). "Defining and differentiating mathematical learning disabilities and difficulties," in *Why is Math So Hard for Some Children: The Nature and Origins of Mathematics Learning Difficulties and Disabilities,* eds D. B. Berch and M. M. M. Mazzocco (Baltimore, MD: Paul H. Brookes), 29–47.

Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Dev.* 82, 1224–1237. doi: 10.1111/j.1467-8624.2011.01608.x

Mazzocco, M. M., and Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Ann. Dyslexia* 53, 218–253. doi: 10.1007/s11881-003-0011-7

Mazzone, L., Ducci, F., Scoto, M. C., Passaniti, E., Genitori D'Arrigo, V., and Vitiello, B. (2007). The role of anxiety symptoms in school performance in a community based sample of children and adolescents. *BMC Public Health* 7:347. doi: 10.1186/1471-2458-7-347

Moeller, K., Fischer, U., Link, T., Wasner, M., Huber, S., Cress, U., et al. (2012). Learning and development of embodied numerosity. *Cogn. Process.* 13(Suppl. 1), 271–274. doi: 10.1007/s10339-012-0457-9

Morsanyi, K., Devine, A., Nobes, A., and Szucs, D. (2013). The link between logics, mathematics and imagination: evidence from children with developmental dyscalculia and mathematically gifted children. *Dev. Sci.* 16, 542–553. doi: 10.1111/desc.12048

Moser Opitz, E., and Ramseier, E. (2012). Rechenschwach oder nicht rechenschwach. Eine kritische Auseinandersetzung mit Diagnosekonzepten, Klassifikationssystemen und Diagnoseinstrumenten unter besonderer Berücksichtigung von älteren Schülerinnen und Schülern [Mathematical Learning Disabilities: a critical examination of diagnostic conceptx, classification systems and diagnostic instruments, with particular consideration of older students]. *Lern. Lernstör.* 1, 99–117. (Full article in German, extended abstract in English).

Murphy, M. M., Mazzocco, M. M., Hanich, L. B., and Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *J. Learn. Disabil.* 40, 458–478. doi: 10.1177/00222194070400050901

National Center for Education Statistics. (2011). *The Condition of Education 2011.* Retrieved from http://nces.ed.gov/pubs2011/2011033.pdf

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012

Pixner, S., and Kaufmann, L. (2013). Prüfungsangst, Schulleistung und Lebensqualität bei Schülern [Exam anxiety, school achievement and quality of life in third and sixth grade students]. *Lern. Lernstör.* 2, 111–124. (Full article in German, extended abstract in English). doi: 10.1024/2235-0977/a000034

Price, G. R., Mazzocco, M. M., and Ansari, D. (2013). Why mental arithmetic counts: brain activation during single digit arithmetic predicts high school math scores. *J. Neurosci.* 33, 156–116. doi: 10.1523/JNEUROSCI.2936-12.2013

Rourke, B. P., and Conway, J. A. (1997). Disabilities of arithmetic and mathematical reasoning perspectives from neurology and neuropsychology. *J. Learn. Disabil.* 30, 34–46. doi: 10.1177/002221949703000103

Rubinsten, O., Bedard, A.-C., and Tannock, R. (2008). Methylphenidate has differential effects on numerical abilities in ADHD children with and without co-morbid mathematical disabilities. *Open Psychol. J.* 1, 11–17. doi: 10.2174/1874350100801010011

Rubinsten, O., and Henik, A. (2005). Automatic activation of internal magnitudes: a study of developmental dyscalculia. *Neuropsychology* 19, 641–648. doi: 10.1037/0894-4105.19.5.641

Rubinsten, O., and Henik, A. (2009). Developmental dyscalculia: heterogeneity might not mean different mechanisms. *Trends Cogn. Sci.* 13, 92–99. doi: 10.1016/j.tics.2008.11.002

Rubinsten, O., Henik, A., Berger, A., and Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *J. Exp. Child Psychol.* 81, 74–92. doi: 10.1006/jecp.2001.2645

Shalev, R. S., and Gross-Tsur, V. (1993). Developmental dyscalculia and medical assessment. *J. Learn. Disabil.* 36, 134–137. doi: 10.1177/002221949302600206

Shalev, R. S., Manor, O., Kerem, B., Ayali, M., Badihi, N., Friedlander, Y., et al. (2001). Familial-genetic facets of developmental dyscalculia. *J. Learn. Disabil.* 34, 59–65. doi: 10.1177/002221940103400105

Siegler, R. S. (1987). The perils of averaging data over strategies: an example from children's addition. *J. Exp. Psychol. Gen.* 116, 250–264. doi: 10.1037/0096-3445.116.3.250

Siegler, R. S. (1995). How does change occur: a microgenetic study of number conservation. *Cogn. Psychol.* 28, 225–273. doi: 10.1006/cogp.1995.1006

Stock, P., Desoete, A., and Roeyers, H. (2010). Detecting children with arithmetic disabilities from kindergarten: evidence from a 3-year longitudinal study on the role of preparatory arithmetic abilities. *J. Learn. Disabil.* 43, 250–268. doi: 10.1177/0022219409345011

Temple, C. M. (1991). Procedural dyscalculia and number fact dyscalculia: double dissociation in developmental dyscalculia. *Cogn. Neuropsychol.* 8, 155–176. doi: 10.1080/02643299108253370

Vogel, S. E., and Ansari, D. (2012). Neurokognitive Grundlagen der typischen und atypischen Zahlenverarbeitung. [Neurocognitive foundations of typical and atypical number development]. *Lern. Lernstör.* 1, 135–149. (Full article in German, extended abstract in English). doi: 10.1024/2235-0977/a000015

von Aster, M. G. (2000). Developmental cognitive neuropsychology of number processing and calculation: varieties of developmental dyscalculia. *Eur. Child Adolesc. Psychiatry* 9, 41–58. doi: 10.1007/s007870070008

von Aster, M. G., and Shalev, R. S. (2007). Number development and developmental dyscalculia. *Dev. Med. Child Neurol.* 49, 868–873. doi: 10.1111/j.1469-8749.2007.00868.x

Wilson, A. J., and Dehaene, S. (2007). "Number sense and developmental dyscalculia," in *Human Behavior and the Developing Brain*, eds D. Coch, K. Fischer, and G. Dawson (New York, NY: Guildford Press), 212–238.

Design and evaluation of the computer-based training program Calcularis for enhancing numerical cognition

Tanja Käser[1]*, Gian-Marco Baschera[1], Juliane Kohn[2], Karin Kucian[3,4], Verena Richtmann[2], Ursina Grond[3,4], Markus Gross[1] and Michael von Aster[3,4,5]

[1] Department of Computer Science, ETH Zurich, Zurich, Switzerland
[2] Department of Psychology, University of Potsdam, Potsdam, Germany
[3] Center for MR-Research, University Children's Hospital, Zurich, Switzerland
[4] Children's Research Center, University Children's Hospital, Zurich, Switzerland
[5] Department of Child and Adolescent Psychiatry, DRK Kliniken Berlin Westend, Berlin, Germany

This article presents the design and a first pilot evaluation of the computer-based training program Calcularis for children with developmental dyscalculia (DD) or difficulties in learning mathematics. The program has been designed according to insights on the typical and atypical development of mathematical abilities. The learning process is supported through multimodal cues, which encode different properties of numbers. To offer optimal learning conditions, a user model completes the program and allows flexible adaptation to a child's individual learning and knowledge profile. Thirty-two children with difficulties in learning mathematics completed the 6–12-weeks computer training. The children played the game for 20 min per day for 5 days a week. The training effects were evaluated using neuropsychological tests. Generally, children benefited significantly from the training regarding number representation and arithmetic operations. Furthermore, children liked to play with the program and reported that the training improved their mathematical abilities.

Keywords: learning, intervention, optimization, calculation, spatial representation, interactive learning environment

INTRODUCTION

Arithmetical skills are essential in modern society. However, many children experience difficulties in learning mathematics, ranging from mild to severe numeracy problems. It is therefore important to investigate the typical and atypical development of mathematical abilities as well as intervention approaches to prevent or remediate difficulties. In this study, we present the development of a computer-based training program for children with difficulties in learning mathematics along with case studies and quantitative results of a first evaluation.

In the following, we first introduce different neuro-cognitive models of number processing and numerical development focusing on the models relevant for the design of the training program. We then discuss the potential of computer-based training environments and give an overview of existing interventions before introducing the present study.

NEURO-COGNITIVE MODELS OF NUMBER PROCESSING AND NUMERICAL DEVELOPMENT

Current neuropsychological models postulate distinct representational modules, located in different brain areas, which are relevant for adult cognitive number processing and calculation. One of the first models, the "triple-code model" (Dehaene and Cohen, 1995) comprises a verbal module supporting counting and number fact retrieval, a visual-Arabic module required for solving written arithmetic and an analogue magnitude module (mental number line) for semantic number processing. Lately, an fMRI meta-analysis enabled further insights into supporting and domain-general functions involved in solving arithmetic tasks and suggested a modification and extension of the triple-code model (Arsalidou and Taylor, 2011). Results from functional brain imaging in adults and children indicate that the representation of the mental number line emerges during the first years of school in the parietal lobe due to practice and experiences (Rivera et al., 2005; Ansari and Dhital, 2006; Kucian et al., 2008). The initial assumption of the analogue magnitude representation being notation-independent was challenged in 2007 (Cohen Kadosh et al., 2007). Nieder (2012) recently showed that there are indeed notation-dependent as well as notation-independent neurons responding to numerosity.

While the triple-code model denotes the end state of numerical development, the four-step developmental model (von Aster and Shalev, 2007) describes the path to this end state. It divides the semantic representation (analogue magnitude representation) into an implicit core representation of magnitude and an explicit mental number line, the latter considered as being a "representational redescription" of the former (Karmiloff-Smith, 1992). The (inherited) core-system representation of cardinal magnitude provides the basic meaning of numbers (step 1). Based on this representation, children learn to associate a perceived number with spoken and later written and Arabic symbols. The process of linguistic (step 2) and Arabic (step 3) symbolization is in turn a precondition for the development of a mental number line (step 4). The different representations develop depending on the growing capacity of domain-general functions like working memory.

Lately, other authors have suggested different models of numerical development (Carey, 2001, 2004; Kucian and Kaufmann, 2009; Kaufmann et al., 2011; Noel and Rousselle, 2011; Kaufmann and von Aster, 2012; Vogel and Ansari, 2012). Some authors argue that developmental dyscalculia (DD) is mainly caused by an early, probably genetic, deficit of the basic non-symbolic magnitude system (Butterworth et al., 2011), while others suggest that problems may arise from different developmental reasons, including maladaptive learning experiences and math anxiety (see also the opinion paper Kaufmann et al., submitted). To summarize, there is still an open debate about developmental trajectories and reasons for failure in learning mathematics. However, there seems to be agreement that based on early non-symbolic abilities to access and compare numerical magnitudes, different components of semantic and symbolic representations are developing during childhood and school years. These components develop based on the increasing capacity of domain-general functions and enable a child to successively acquire arithmetic skills.

COMPUTER-BASED INTERVENTIONS

The highly complex processes of domain-specific cognitive development need to be taken into account when teaching mathematics. The development of each child's numerical abilities often follows a different speed and is intertwined with the development of other cognitive domains and domain-general abilities (von Aster and Shalev, 2007; Kucian and Kaufmann, 2009; Kaufmann et al., 2011), leading to different mathematical performance profiles (von Aster, 2000; Geary, 2004; Wilson and Dehaene, 2007). Therefore, a high grade of individualization seems necessary.

Educational software can contribute to these requirements. Computer-based trainings can be designed to adapt to an individual child's abilities and provide intensive training in a stimulating environment (Kullik, 2004). The training can for example adapt to cognitive (Naglieri and Johnson, 2000) or to performance profiles of the children (von Aster, 2000; Geary, 2004; Wilson and Dehaene, 2007). This individualization in combination with the fact that the computer is an emotionally neutral medium may also lead to increased motivation and enhance positive self-concepts as every learner gains feelings of success (Ashcraft and Faust, 1994; Spitzer, 2009). Furthermore, computers are an attractive medium for children (Kulik and Kulik, 1991; Schoppek and Tulis, 2010).

In the past years, different meta-analyses have assessed the effects of computer-based instruction, revealing positive results. Kulik and Kulik (1991) and Kulik (1994) computed an average effect size of 0.47 for math learning in elementary school. Other studies reported effect sizes ranging from 0.13 to 0.8 (Khalili and Shashaani, 1994; Fletcher-Flinn and Gravatt, 1995). Li and Ma (2010) found larger effects for elementary school than for higher education and showed that special needs students especially benefit from computer-based instruction.

Existing interventions are, however, mostly conventional. Techniques include training programs for preschool children at risk of developing mathematical difficulties (Griffin et al., 1994; Van De Rijt and Van Luit, 1998; Arnold et al., 2002; Wright, 2003) as well as remedial programs for elementary school children (Van Luit and Naglieri, 1999; Dowker, 2001, 2003; Fuchs et al.,

2006; Wilson et al., 2006a; Butterworth et al., 2011; Kucian et al., 2011; Lenhard et al., 2011). Programs designed for preschool children mostly focus on building basic-numerical skills, whereas elementary school trainings target a broader range of skills. Some interventions address basic numerical skills and the establishment of the mental number line (Wilson et al., 2006a), while others train arithmetic fact knowledge (Van Luit and Naglieri, 1999; Fuchs et al., 2006) or are aligned to scholar curricula (Lenhard et al., 2011). Other effective approaches combine the training of basic-numerical capacities with the training of arithmetical knowledge (Dowker, 2001, 2003; Kucian et al., 2011).

The computer-based intervention "Number race" for children with DD trains number comparisons and enhances the links between number and space (Wilson et al., 2006a). Evaluation of the training revealed significant improvements in basic numerical cognition, but the effects did not generalize to counting or arithmetic (Wilson et al., 2006b, 2009; Räsänen et al., 2009). "Rescue Calcularis" is another computer-based intervention for children with DD. It aims to improve the construction and access to the mental number line. The evaluation of the program showed that children with and without DD could benefit from the training (Kucian et al., 2011). "Elfe and Mathis" is a computer-based training (Lenhard et al., 2011) aligned to the German scholar curriculum. Its evaluation demonstrated significant effects. Fuchs et al. (2006) presented a computer-based program to acquire fact knowledge, reporting significant effects in addition. Butterworth et al. (2011) suggest the use of adaptive interactive games for remediation (see also Callaway, 2013). The proposed games train basic-numerical skills (number comparisons and counting) as well as the spatial number representation and simple arithmetic facts. Evaluative results of the training have not yet been published.

The previous studies demonstrate the efficacy of computer-based intervention in number processing. The presented programs however mostly focus on specific skills (such as the training of fact knowledge) and provide only limited adaptability.

THE PRESENT STUDY

The objective of the present study is (1) the development of a computer-based training program based on current concepts of numerical development and (2) a first pilot evaluation of its efficacy and practicality. The intervention uses core elements of "Rescue Calcularis" (Kucian et al., 2011). Compared to previous studies, we provide a more complete training of mathematical skills and employ a user model allowing flexible adaptation [based on the student model and control algorithm presented in Käser et al. (2012)]. Our program combines the training of basic numerical cognition with the training of arithmetical abilities. Several past studies (Siegler and Booth, 2004; Booth and Siegler, 2006, 2008; Halberda et al., 2008) have reported significant correlations between math achievement and arithmetical learning and the quality of numerical magnitude representation. In addition, intervention programs that train basic numerical skills and arithmetic knowledge in parallel have proven to be successful (Dowker, 2001, 2003; Kucian et al., 2011). Based on these facts, we expect significant training effects regarding spatial number representation as well as arithmetic performance. Furthermore, we expect

an increased motivation by providing an attractive computer-based learning environment and by adapting the difficulty level to the individual child.

METHODS

DESCRIPTION OF THE TRAINING PROGRAM

The training program combines the training of basic numerical cognition with the training of different number representations and their interrelations and arithmetical abilities. Our intervention relies on three design principles:

(1) Design of numerical stimuli: A special number design enhancing the different number representations is consistently used throughout the training program. Furthermore, the three different number modalities are shown simultaneously at the end of each trial.
(2) Adaptability and scaffolding: The learning process of children is different, i.e. they start at different initial levels, progress with different speed and show different specific problems. Our intervention program optimizes the learning process by providing a hierarchically structured learning environment teaching fundamental knowledge first (scaffolding). Furthermore, it is fully adaptive to customize the learning experience of the child. Task difficulty is adapted to the child and specific problems of a child are recognized and addressed.
(3) Different types of knowledge: The intervention program aims to balance the acquisition of conceptual knowledge with automation training. Children are taught conceptual knowledge before going over to automation training. An arithmetic operation is for example first introduced and explained. The arithmetic operation and its solution are then modeled using stimuli and finally, mental calculation is trained.

Design for numerical stimuli

The special design for numerical stimuli is intended to enhance the different number modalities and to strengthen the links between them. Properties of numbers are encoded with visual cues such as color, form and topology. The digits of a number are attached to the branches of a graph and represented with different colors according to their positions in the place-value system: Units are colored in green, tens in blue and hundreds in red (**Figure 1** left). We assume that this representation facilitates the acquisition of the Arabic notation as well as the translation between verbal and Arabic notation. The cardinal magnitude of number is emphasized by representing the number as an assembly of one, ten and hundred blocks. This representation illustrates the fact

FIGURE 1 | Numerical stimuli for the number 35. Number graph (left), colored block (middle) and number line representation with integrated blocks (right).

that numbers are composed of other numbers. The blocks are linearly arranged from left to right (**Figure 1** middle) or directly integrated in the number line (**Figure 1** right).

Structure

The training program is composed of multiple games in a hierarchical structure. **Figure 2** shows the target structure of the training program. The study version (section User study) of the program is constrained to specific areas of the target structure (intuitive number understanding, number representations, arithmetic operations) and to natural numbers up to 1000. In the following, we describe the target structure of the training program.

Games are structured along number ranges and further divided into hierarchically ordered areas:

(1) Number representations: This area focuses on different number modalities and number understanding in general. It trains transcoding between different number representations. Furthermore, the three interpretations of number are established: Cardinality (quantity), ordinality (position in a sequence) and relativity (difference between two numbers). Games in this area are hierarchically ordered according to the four-step developmental model (von Aster and Shalev, 2007).
(2) Arithmetic operations: This area trains arithmetic operations at different difficulty levels. Task difficulty is determined by task complexity, the magnitude of numbers involved and the means (visual aids) available to solve the task.
(3) Word problems: A complete understanding of mathematical operations requires the ability to associate a described situation with a mathematical operation and vice-versa. This also presumes an understanding of the actual meaning of the operation.

Each area builds up on knowledge gained in previous areas and implicitly trains these skills further. An additional forth area serves as a precondition for the three areas described above. This area focuses on important precursor abilities (Landerl et al., 2004; Hannula and Lehtinen, 2005; Mazzocco and Thompson, 2005; Krajewski and Schneider, 2009) such as subitizing or counting.

Games can also be categorized based on their complexity and relative importance. Main games are complex games requiring a combination of abilities to solve them. Support games train specific skills and serve as a prerequisite for the main games. Each area features one main game and several support games. A typical training path would traverse each number range from left to right starting with the number range from 0 to 10. The main games are the same for each number range; they just differ by the cardinal magnitude of numbers used.
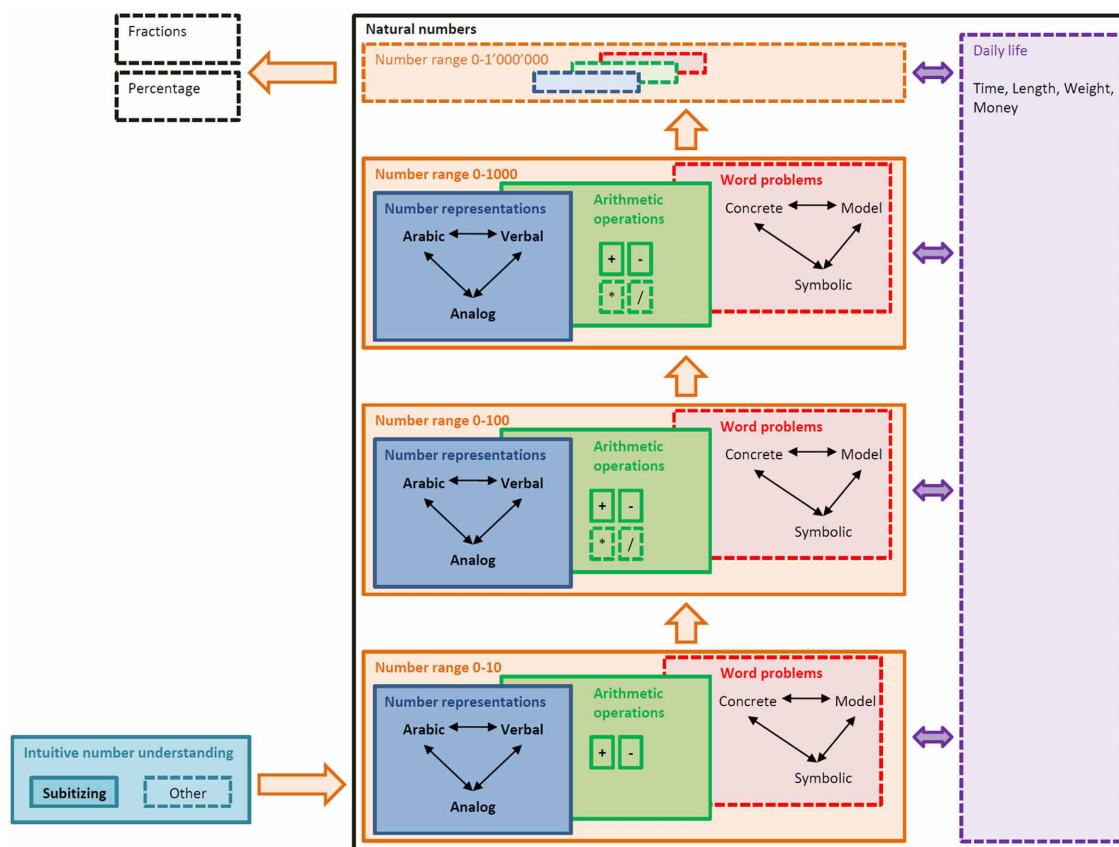
Adaptive algorithm

To offer optimal learning conditions, the training program adapts to the needs of a specific child. All children start the training with the same game. After each trial, the program estimates the actual knowledge state of the child and displays a new task adjusted to this state. The student model and controller mechanism are based
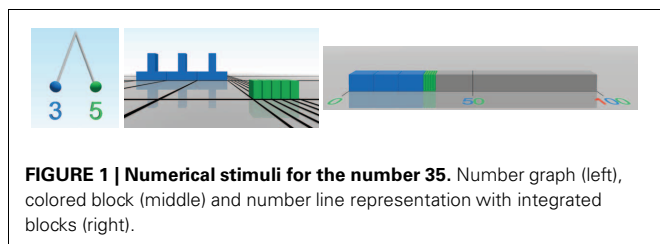
FIGURE 2 | Target structure of the training program. The continuous lines mark the areas present in the study version.

on the mathematical concepts presented by Käser et al. (2012) and were developed for the study version of the training structure.

Student model. We model the mathematical knowledge using a dynamic Bayes net. This net consists of a directed acyclic graph representing different mathematical skills and their relationships. The skills are connected based on the dependencies among them, i.e., two abilities A and B have a (directed) connection, if having ability A is a precondition for having ability B. As the skills cannot be observed directly, the program infers them by posing specific tasks and evaluating user actions. Therefore, we assign all types of tasks and their outcome to the different skills. The resulting student model contains 100 different skills. The skills can be assigned to the different areas of the training program.

Figure 3A displays the skills of the area "Number representations" in the number range from 0 to 100. The skills colored in blue denote the different number representations: *Concrete* (number as a set of objects), *Verbal* (spoken number), *Arabic* (written number) and *Numberline* (number as a position on a number line). The transcoding skills (translation between two number representations) are colored in red. The yellow skills introduce the principles of ordinality and relativity of number. Children are required to give the precursor or successor of a number (*Ordinal 1*) or to add/subtract 10 (or 20 or 30) from a given
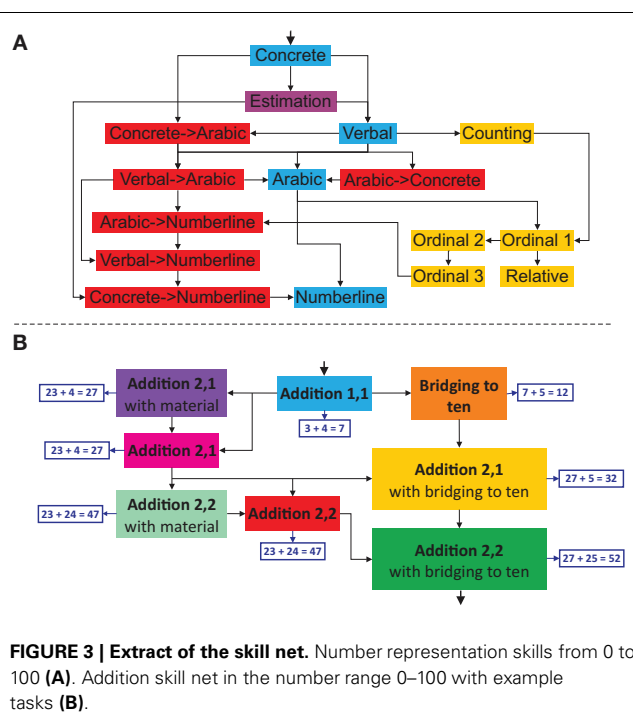
FIGURE 3 | Extract of the skill net. Number representation skills from 0 to 100 **(A)**. Addition skill net in the number range 0–100 with example tasks **(B)**.

number (*Relative*). In another task, numbers need to be ordered according to their magnitude (*Ordinal 2*). Finally, children need to guess a number in the range from 0 to 100 (*Ordinal 3*). The purple skill trains estimation, i.e., children are required to estimate the quantity of a given point set. Skills in this area are hierarchically ordered according to the four-step developmental model (von Aster and Shalev, 2007). Following the model, the transcoding between the linguistic and Arabic symbolization (*Verbal->Arabic*) is trained before giving the position of a written number on a number line (*Arabic->Numberline*).

The skills in the area "Arithmetic operations" are ordered according to their difficulty. **Figure 3B** displays the addition skills in the number range from 0 to 100. The difficulty of a task depends on the magnitude of the involved numbers, its complexity and the means allowed solving it. Computing $23 + 24 = 47$ (*Addition 2,2*) is considered more difficult than calculating $3 + 4 = 7$ (*Addition 1,1*) as the latter task involves smaller numbers. Furthermore, a task involving bridging to 10 such as $27 + 5 = 32$ (*Addition 2,1 with bridging to 10*) is rated more complex than a task without any crossing. And finally, modeling the task $23 + 4 = 27$ (*Addition 2,1 with material*) is easier than calculating the task mentally (*Addition 2,1*). Subtraction skills in the number range 0 to 100 exhibit exactly the same structure as the addition skills. Therefore, one can obtain the skill net for subtraction by simply replacing *Addition* by *Subtraction* in **Figure 3B**.

Each skill has two states: A learnt state and an unlearnt state. Having a dynamic Bayes net, the probability for a skill being in the learnt state can be computed. All probabilities are initialized to 0.5, as the system does not know anything about the knowledge of the child. The probabilities are updated after each trial. The probability of a skill can be influenced in different ways. On the one hand, it changes, if the child solves a task that is associated with this skill. On the other hand, solving a task that is associated with a precursor or a successor skill influences the probability.

Controller. The game controller selects the skills for training. After each child input, the controller selects one of the following options based on the probability of the current skill:

(1) Stay: Continue the training of the current skill.
(2) Go back: Train a precursor skill.
(3) Go forward: Train a successor skill.

The decision is based on an upper and lower border for the probability of the current skill. If the probability is larger than the upper border, a more difficult skill is selected for training. If it is smaller than the lower border, a precursor skill is selected for training. The area between the borders is considered as being optimal for training. The two borders have been chosen heuristically to reach the desired behavior: in order to pass a skill, about 10 tasks in a row need to be solved correctly. About five tasks in a row lead to failing a skill.

As a skill can have multiple precursor or successor skills, there are several options for going back or forward. The basic assumption of the model is that in order to pass the current skill, all the precursor skills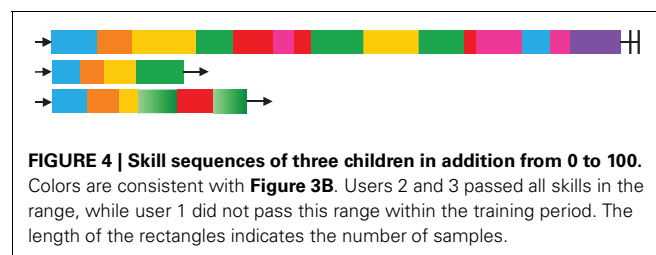 need to be mastered. If the child therefore fails a skill, the controller selects a precursor skill that has not yet been played for training. When going forward, the control algorithm prefers main games over support games (section Structure) and thus chooses the shortest way through the skill net. A detailed description of the mathematical model and control algorithm can be found in Käser et al. (2012).

At the beginning of the training, children start with the lowest skill in each area. Due to the structure of the skill net and the control algorithm acting on it, each child persecutes a different trajectory through the skill net during training. This variety is increased by repeating less sophisticated skills at random time intervals. Therefore, the path through the network is different for every user (see **Figure 4**).

To allow an even more accurate adaptation, the program has access to a bug library storing typical error patterns (Gerster, 1982). If a child commits a typical error several times, the controller systematically selects actions for remediation. **Table 1** lists the typical error patterns stored in the bug library, along with examples and remediation tasks. For the area of number representations, only one pattern is stored for the landing game: positioning the cone on the wrong side of the indicated center of the number line, i.e., positioning the cone at a number <50 when the given number is >50. For the area of arithmetic operations, a range of error patterns are stored in the bug library. Some of these patterns can be attributed to problems in counting or understanding the basic concepts of addition and subtraction. Remediation skills for these error patterns train simple addition and subtraction tasks with colored blocks (*Addition/Subtraction 1,1 with material,* "slide rule" game). Other error patterns probably occur due to a lack in understanding the Arabic notation system, i.e., the meaning of the different positions of the digits. Selected remediation action for these patterns is the training of the Arabic notation system (*Arabic->Concrete*). Another typical error is the switching of digits (twenty-five is written as "52") which is remediated by training transcoding from spoken to written numbers (*Verbal->Arabic*). Finally, problems with bridging to 10 are also addressed (*Bridging to ten*). The bug library was built based on previous work identifying typical error patterns and their causes (Gerster, 1982). In a next step, the typical error patterns will be analysed and refined based on the collected input data.

Games

The training program consists of 10 different types of games that are associated with the presented skills. By varying the numbers used in the games, we obtain 81 different types of tasks (task difficulty levels). In the following, we describe four games of the training program.

FIGURE 4 | Skill sequences of three children in addition from 0 to 100. Colors are consistent with **Figure 3B**. Users 2 and 3 passed all skills in the range, while user 1 did not pass this range within the training period. The length of the rectangles indicates the number of samples.

Table 1 | Description of typical errors along with examples and remediation games for the domains of number representations (NR), addition (A) and subtraction (S).

	Description	Example	Remediation
NR	Landing game: the child lands the cone on the wrong side of the center (5, 50, or 500).	–	Training of the ordering of numbers according to their magnitude (*Ordinal 1*)
A, S	Correct result is missed by 1 (±1).	$5 + 3 = 7$	Training of addition or subtraction with colored blocks (*Addition/Subtraction 1,1 with material*)
A, S	Addition instead of subtraction (or vice versa).	$5 + 3 = 2$	Training of addition or subtraction with colored blocks (*Addition/Subtraction 1,1 with material*)
A	Addition of all digits.	$12 + 24 = 9$	Training of the Arabic notation system (*Arabic->Concrete*)
A, S	Switching of digits when reading/writing a number.	$24 - 3 = 12$	Transcoding from spoken to written notation (*Verbal->Arabic*)
A, S	Use of wrong digit order.	$63 - 5 = 13$	Training of the Arabic notation system (*Arabic->Concrete*)
A, S	Forgetting the carry when bridging to ten.	$34 + 7 = 31$	Training of bridging to 10 using colored blocks (*Bridging to ten*)
A, S	Addition/Subtraction of inner and outer digits.	$34 + 13 = 56$	Training of the Arabic notation system (*Arabic->Concrete*)
S	Building the difference between digits.	$34 - 17 = 23$	Training of the Arabic notation system (*Arabic->Concrete*)

Ordering. The "ordering" game (**Figure 5A**) is a support game in the area of "Number Representations," training ordinal number understanding. A sequence of numbers is displayed for a period of 5 s. Children need to decide, if the sequence was sorted in ascending order. The game is associated with the skill *Ordinal 1* in **Figure 3A**.

Landing. The "landing" game (**Figure 5B**) is the main game in the area of "Number Representations," training spatial number representation. A purple cone must be directed to the position of a given number on a number line (with indicated center), using a joystick. Numbers are presented in verbal or Arabic notation. In another option the cardinality of a given point set and the position of this quantity on the number line have to be estimated. The different modes of the game are associated with the skills *Verbal->Numberline*, *Arabic->Numberline* and *Concrete->Numberline* in **Figure 3A**. The required accuracy for a correct solution is a deviance of less than 5%.
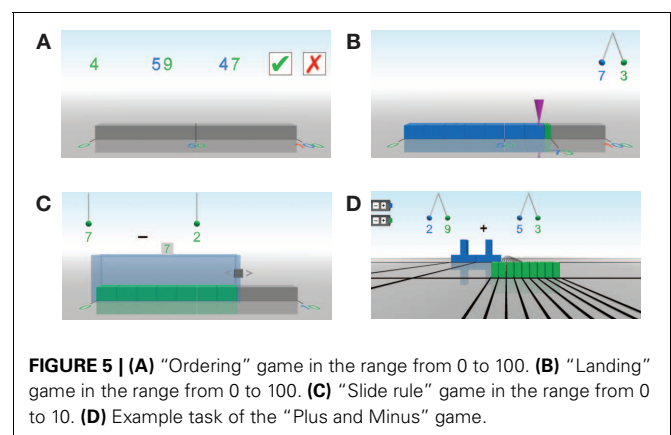
Slide rule. The "slide rule" game (**Figure 5C**) is a support game belonging to the area of "Arithmetic operations," providing an introduction to addition and subtraction using the part-whole concept. An operation task is presented to the child, as well as a number line and a glass case containing a number of unit blocks (according to the first number of the task). The size of the glass case must be changed such that it contains the result of the task. This game would be associated with the skills *Addition 1,1 with material* and *Subtraction 1,1 with material*.

Plus and Minus. The "Plus and Minus" (**Figure 5D**) game is a support game in the area of "Arithmetic operations." An arithmetic operation given in Arabic notation must be modeled using colored blocks (one, ten, and hundred). Different strategies are allowed to find the result. This game is associated with all addition and subtraction skills that involve the use of materials.

USER STUDY
Study design and participants
The effects of the training program have been assessed in a study with 41 children conducted in Switzerland. Participants were

FIGURE 5 | (A) "Ordering" game in the range from 0 to 100. **(B)** "Landing" game in the range from 0 to 100. **(C)** "Slide rule" game in the range from 0 to 10. **(D)** Example task of the "Plus and Minus" game.

divided into a training group ($n = 20$, 65% females) completing a 12-weeks training and a waiting group ($n = 21$, 66.6% females) starting with a 6-weeks rest period. Comparing the training effects of the training group to those of a waiting group allows controlling for developmental and schooling effects.

Mathematical performance of both groups was evaluated at the beginning of the study (t_1), after 6 weeks (t_2) and after 12 weeks (t_3). Children were required to train with the program 5 times per week, with daily training sessions of 20 min. The groups were matched according to age (training group: $M = 9.96$ years ($SD = 1.35$), min $= 7.37$, max $= 12.06$; waiting group: $M = 9.98$ ($SD = 1.33$), min $= 7.52$, max $= 12.21$; $t_{(39)} = -0.04$, $p = 0.96$), gender and intelligence (training group CFT-score: $M = 93.8$ ($SD = 11.9$); waiting group CFT-score: $M = 93.5$ ($SD = 14.1$); $t_{(39)} = 0.07$, $p = 0.95$) (Cattell et al., 1997; Weiss, 2006). Groups were built by forming matched pairs of kids, followed by a quasi-random assignment to either the training or waiting group (ensuring that the number of males was balanced between the groups).

All participants were German-speaking and visited the 2nd–5th grade of elementary school. Children were indicated by parents and teachers as exhibiting difficulties in learning mathematic. On average, arithmetic performance [measured with the "Heidelberger Rechentest" HRT (Haffner et al., 2005)] of the

participants was around the 10th percentile, corresponding to a T-score of 37 [HRT addition T-score: $M = 37.15$ ($SD = 7.69$); HRT subtraction T-score: $M = 37.29$ ($SD = 8.77$)]. There was no significant difference in arithmetic performance between the groups (HRT addition: $t_{(39)} = 0.59$, $p = 0.55$; HRT subtraction: $t_{(39)} = -0.63$, $p = 0.53$).

Children performed the training at home with exception of one mandatory training session per 6 weeks at our laboratory. Children received a sticker per completed training session that they could put on their training progress sheet. During the training period, all the input data of the children was saved. Therefore, the exact training time of the children could be determined at the end of the study and children with an insufficient number of sessions were excluded from the analysis (see section Results). Parents gave informed consent and children received a small gift for their participation. The presented evaluation was a first pilot study conducted in the context of a large-scale multi-center evaluation study in Germany and Switzerland, which was approved by the ethics committee of the University of Potsdam.

Instruments

All children underwent a series of mathematical performance and number processing tests, detailed below. The children completed a questionnaire after the training, including questions on difficulty, motivation, and personal evaluation of the training.

Heidelberger Rechentest (HRT). Arithmetic performance was assessed using the addition and subtraction subtests of the HRT (re-test reliability: addition $r_{tt} = 0.82$, subtraction $r_{tt} = 0.86$). In these subtests, children are presented a list of addition (subtraction) tasks ordered by difficulty. The goal is to solve as many tasks as possible within 2 min. The maximum number of correct tasks is 40. During the test sessions, the addition subtest of the HRT was always solved first, followed by the subtraction subtests and the computer-based tests described below.

Computer-based tests. Children also underwent a series of computer-based mathematical tests (see **Figure 6**):

- Arithmetic (AC) (**Figure 6A**): In this test, children solve a series of addition (subtraction) tasks. Trials are ordered by difficulty and presented serially. The time to solve the tasks is 10 min. The maximum number of solved tasks is 76.
- Number line (NL) (**Figure 6B**): In this test, children need to indicate the position of a given number (presented in Arabic

notation as well as verbally) on a number line. The number line is represented on the screen as a one-dimensional black line with labeled end points. The position of the number can be indicated by mouse-click. There are 10 tasks in the number range from 0 to 10 (NL 10), 20 tasks between 0 and 100 (NL 100) and 10 tasks between 0 and 1000 (NL 1000).
- Non-symbolic magnitude comparison (NC) (**Figure 6C**): In this test, children are presented 10 sets with 1–9 black dots (excluding 5) for a period of 120 ms. Children need to indicate if the presented number of dots was smaller or larger than 5. The representation of the dots is balanced according to spatial distribution and area properties as described by Rubinsten and Sury (2011). The black area is the same for all trials. Half of the trials have a small extension (high density) while the other half is spread out (low density).
- Estimation: In this task, children are presented twenty sets with 1–99 black dots (excluding 50). Children need to decide, if the presented sets are smaller or larger than 50. Numbers are equally distributed over the range. Confounding visual factors are controlled as described in the non-symbolic magnitude comparison task. Stimuli are shown for a period of 5 s.

During the test sessions, the different tests were solved in the following order: AC addition, NL 0–10, NC, AC subtraction, NL 0–100, estimation, NL 0–1000. The computer-based tests exist in three parallelized versions (one per measurement point). The versions were parallelized according to content and item difficulty. Each version of the addition and subtraction tests for example contains the same number of tasks between 0 and 10 and the same number of tasks involving bridging to 10.

Feedback questionnaire. Children completed a training evaluation questionnaire at the end of the study (t_3). Children indicated for each game, how much they liked it. The scale was represented through smileys, going from a laughing (4) to a crying (0) smiley. The difficulty of the training was judged on a scale from very easy (0) to very difficult (4). And finally, children needed to indicate if the training helped them on a scale from not true (0) to absolutely correct (3).

RESULTS

Only children with at least 24 sessions after the 6-weeks training period were included in the evaluation of the training. Thus, five children from the training group (4: technical challenges, 1: <24 training sessions) and four children from the waiting group

FIGURE 6 | (A) Addition task. **(B)** Number line task between 0 and 10. **(C)** Estimation task.

(1: abort of study, 3: <24 training sessions) were excluded from the analysis. The exclusions did not change the matching of the groups. **Table 2** gives an overview of the training statistics.

QUANTITATIVE ANALYSES

A repeated measures general linear model (GLM) analysis was conducted to evaluate training effects (t_1–t_2) as a within-subject factor and group (Training/Waiting) as a between-subject factor. *Post-hoc* paired-sample t-tests were used to test for differences in performance for consecutive testing periods ($t_1 - t_2$, $t_2 - t_3$). Effect sizes were computed according to Field (2009). No corrections for multiple testing were applied. **Table 3** summarizes the means and standard deviations of the behavioral measures for all measurement points, including calculated statistical results. There were no between-group performance differences prior to the intervention.

Arithmetic (AC addition and subtraction)

The interaction between training and group was significant for subtraction ($p = 0.028$) and showed a trend for addition ($p = 0.081$). Both operations demonstrated medium effect sizes (subtraction: $r = 0.39$, addition: $r = 0.31$). The prolongation of the training from 6 to 12 weeks ($t_2 - t_3$) yielded an additional trend of improvement (addition: $p = 0.072$; subtraction: $p = 0.066$).

HRT (addition and subtraction)

The interaction between training and group was significant only for subtraction (subtraction: $p = 0.002$; addition $p = 0.375$), where children showed a large effect size ($r = 0.52$). The prolongation of the training yielded an additional improvement, which was significant only for addition ($p = 0.004$).

Number line

The quality of the spatial number representation was measured by calculating the distance (percentage) and the variance of the distance between the correct and the indicated location of the number on the number line. In the number range from 0 to 10, children tended to locate the correct position on the number line more accurately after training ($p = 0.058$) and showed decreased variance ($p = 0.022$). The interaction between training and group was significant only for the variance (mean: $p = 0.12$; variance: $p = 0.034$). Children demonstrated medium effect sizes for both measures (mean: $r = 0.28$, variance: $r = 0.38$). The prolongation of the training did not yield any further benefit. In the number

range from 0 to 100, interaction between training and group was not significant (mean: $p = 0.33$; variance: $p = 0.50$). The prolongation of the training had a beneficial effect (mean: $p = 0.042$; variance: $p = 0.05$). In the number range from 0 to 1000, children tended to locate the numbers more accurately only after 12 weeks (mean: $p = 0.096$; variance: $p = 0.331$).

NC and estimation

In these two tasks, the interaction between training and group was not significant (estimation: $p = 0.11$; NC: $p = 0.65$). Unexpectedly, the waiting group showed a significant improvement in the estimation task ($p = 0.039$). This significant result stems from outliers with large improvement (children with 2 correct answers at t_1 and 17 correct answers at t_2) due to not understanding the task at t_1.

Feedback questionnaire

Children generally liked the training [average over all games: $M = 3.0$ ($SD = 0.55$), scale: 0–4] and rated its difficulty as appropriate [$M = 1.7$ ($SD = 0.74$), scale: 0–4]. They also reported that the training helped them to improve in mathematics [$M = 2.1$ ($SD = 0.89$), scale: 0–3].

CASE STUDIES

To illustrate the concept of the learning program and the operation of the controller, the path through the skill net and the training success of a few children is described in the following. The children and their training characteristics are described in **Table 4**. The analyses stem from the 6-weeks training period.

Subtraction 0–100

For subtraction in the range from 0 to 100, the course of training (path through the skill net) has been analyzed for Anne and Jane. **Figure 7** illustrates the sequence of skills of the two children and the respective numbers of samples.

From **Figure 7**, it can be seen, that the path through the skill net is different for each child. While Jane took the straight path through the subtraction section, the path of Anne exhibits several branches as she had to go back and consolidate more basic skills. Furthermore, Jane needed in total only 71 samples to pass the subtraction 0–100 section, whereas Anne solved 241 samples to work through the section. The external training effects in subtraction from 0–100 (measured by the AC subtraction test,

Table 2 | Training statistics [Means (*SD*)] of training group (*n* = 15) and waiting group (*n* = 17).

	6-weeks period (t_1–t_2) Training group	6-weeks period (t_2–t_3) Waiting group	12-weeks period (t_1–t_3) Training group
Number of training sessions	30.2 (3.2)	32.4 (5.2)	49.2 (2.6)
Number of totally solved tasks	1635.0 (293)	1737.0 (266)	2575.0 (414)
Number of solved tasks per session	54.0 (7.2)	54.0 (5.7)	52.2 (7.0)
Highest reached skill [a] (Number representations)	38.6 (8.3)	38.7 (8.4)	40.5 (6.8)
Highest reached skill [a] (Arithmetic operations)	40.5 (14.7)	39.1 (15.5)	43.0 (15.1)

[a] *The skills of the adaptive model are divided into the content areas of the training program (section Adaptive algorithm). Skills in each area are ordered by their number, with the easiest skill having the lowest number.*

Table 3 | Training effects of training group *TG* (*n* = 15) and waiting group *WG* (*n* = 17) on mathematical performance [Means (*SD*)].

Mathematical performance		t_1	t_2	*t*-score (t_2–t_1)	*F*-score[d]	t_3	*t*-score (t_3–t_2)	*ES*[e]
AC Addition[a]	TG	25.9 (10.8)	30.0 (14.1)	2.38*	3.26[+]	34.4 (17.7)	1.95[+]	0.31
	WG	27.4 (10.7)	26.1 (12.0)	−0.56		29.4 (11.7)	1.90[+]	
AC Subtraction[a]	TG	19.2 (12.7)	24.7 (17.1)	2.77*	5.32*	28.8 (17.9)	1.99[+]	0.39
	WG	19.6 (10.2)	18.9 (10.7)	−0.39		26.3 (14.5)	4.11**	
NL10, mean[b]	TG	13.2 (10.1)	9.3 (10.6)	−2.06[+]	2.56	7.5 (5.0)	−0.81	0.28
	WG	10.3 (10.4)	12.3 (11.6)	0.65		6.1 (4.2)	−2.70*	
NL 10, var[c]	TG	10.2 (6.3)	6.9 (6.5)	−2.58*	4.92*	6.7 (4.6)	−0.15	0.38
	WG	7.4 (6.4)	9.4 (7.6)	1.02		4.9 (3.3)	−3.10**	
NL100, mean[b]	TG	10.2 (4.8)	9.6 (6.4)	−0.62	0.98	7.6 (3.3)	−2.24*	0.18
	WG	13.5 (6.0)	11.3 (5.7)	−1.72		9.3 (7.0)	−1.35	
NL100, var[c]	TG	7.7 (3.4)	8.2 (5.1)	0.39	0.47	6.2 (3.0)	−2.15*	0.12
	WG	9.7 (4.3)	9.1 (5.2)	−0.59		8.2 (5.8)	−0.79	
NL1000, mean[b]	TG	18.5 (10.9)	16.1 (7.5)	−1.61	0.70	12.9 (6.7)	−1.79[+]	0.15
	WG	18.0 (7.3)	17.4 (8.1)	−0.44		12.6 (5.6)	−4.20**	
NL 1000, var[c]	TG	13.3 (7.1)	11.9 (5.7)	−0.84	0.00	10.0 (6.4)	−1.01	0.00
	WG	13.5 (5.6)	12.0 (5.7)	−1.13		10.0 (4.4)	−1.87[+]	
Estimation[a]	TG	15.1 (3.9)	14.9 (3.2)	−0.12	2.85	15.8 (2.1)	1.25	0.29
	WG	13.6 (4.5)	16.3 (2.2)	2.25*		15.9 (2.9)	−0.61	
NC[a]	TG	7.9 (1.9)	8.1 (1.5)	0.39	0.21	8.4 (1.6)	0.59	0.08
	WG	7.4 (2.3)	7.8 (2.2)	0.94		8.0 (1.8)	0.19	
HRT Addition[a]	TG	18.7 (5.4)	20.4 (5.6)	2.47*	0.81	22.5 (5.4)	3.46**	0.16
	WG	18.5 (4.8)	19.4 (4.3)	1.27		20.4 (5.7)	1.5	
HRT Subtraction[a]	TG	15.3 (6.1)	19.8 (5.3)	4.85***	11.38**	20.2 (6.2)	0.59	0.52
	WG	16.9 (6.3)	16.9 (5.6)	0.06		18.4 (5.7)	1.5	

[+]*p* < 0.1, *p* < 0.05, **p* < 0.01, ***p* < 0.001.
[a]*Number of correctly solved tasks.*
[b]*Distance (percentage) from correct position.*
[c]*Variance of distance (percentage) from correct position.*
[d]*Time (t_1–t_2) × group.*
[e]*Effect sizes of interaction time (t_1–t_2) × group. r = 0.10: small effect, r = 0.30: medium effect, r = 0.50: large effect.*

Table 4 | Characteristics of the three example cases.

	Sex	Age	Class	Played sessions	Solved tasks	Tasks per session
Anne	Female	8;11	3	28	1272	45.4
Eva	Female	9;8	4	33	1803	54.6
Jane	Female	9;10	4	33	1795	54.4

section Instruments), support this result. In the initial measurement before the training, Jane solved in total 40 tasks, 39 of them correct. She was already proficient in subtraction tasks between 0 and 100 before the training. In contrast, Anne solved in total 26 tasks, 10 of them correct. After the training, Anne managed

to solve 23 tasks correctly; she especially improved in subtraction involving bridging to ten. Also Jane showed an improvement after the training, she solved 49 tasks correctly. However, most of her improvement stems from subtraction tasks in the range from 0 to 1000 (the AC subtraction test contains 32 tasks between 0 and 100; the rest of the tasks is in the range from 0 to 1000).

Number line 0–100
For Eva and Jane, the ability to place a number on a number line (between 0 and 100) was compared. Before the training, Eva managed the task with an average deviation of 11.4 % (measured by the NL 0–100 test, section Instruments). In contrast, Jane reached an average deviation of 5.4%. Thus, Jane was already more accurate than Eva at the beginning of the training. This fact was confirmed during the course of the training. While Eva

needed 127 samples, to pass the landing game (see **Figure 5B**), Jane passed the game with only 21 samples. The maximum deviation for a sample to be rated as correct was 5%. **Figure 8** displays the improvement curves over the course of the training. Recorded input data from all children shows that most samples exhibit an error of 0–20% with only a few samples lying above this range. Therefore, fitting has been done using a generalized linear regression model, assuming a Poisson distribution of the data. The sample indices have been normalized between 0 and 1.

The training sequences of the two children show the same picture. Jane took the direct path through the skill net, whereas Eva had to go back several times. After the training, Eva achieved an average deviation of 6.5% in the NL 0-100 test and Jane's average deviation was 5.1%. Whilst Eva improved significantly, Jane stagnated on a high level.

DISCUSSION

Although many children experience difficulties in learning mathematics, few studies have investigated targeted interventions based on neuro-cognitive findings of the typical and atypical development of mathematical abilities. Only a fraction of these are computer-based. In the present project, we developed a computer-based intervention targeting children with difficulties in learning mathematics and performed a first evaluation. The results achieved are promising and show significant improvements in subtraction and number representation. Moreover, they confirm the behavioral effects obtained in a previous study employing the computer-based training program "Rescue Calcularis" (Kucian et al., 2011).
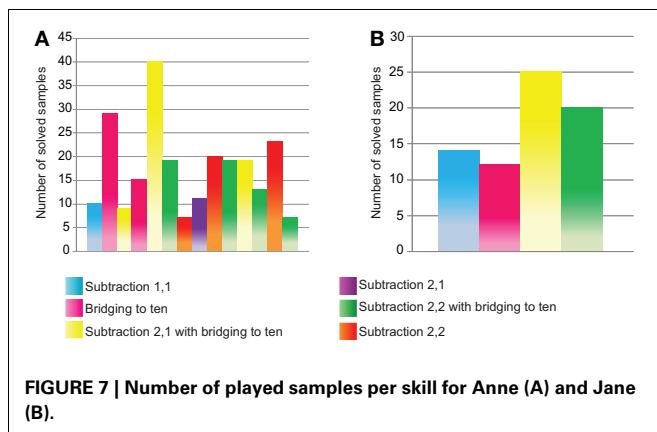
FIGURE 7 | Number of played samples per skill for Anne (A) and Jane (B).
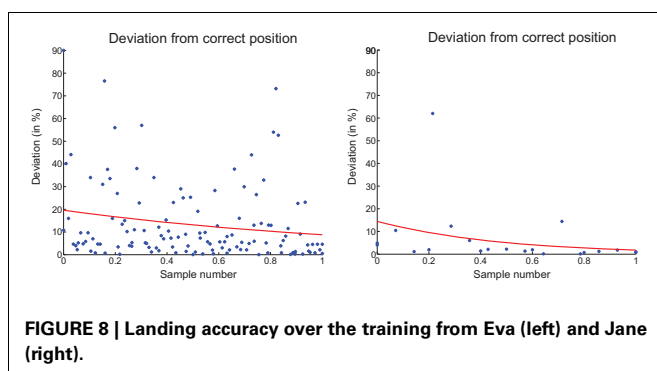
FIGURE 8 | Landing accuracy over the training from Eva (left) and Jane (right).

TRAINING

The first pilot study was conducted not only to assess the efficacy of the training program but also the practicality and adaptability of the learning environment. Feedback from children, who have completed the training and rated the difficulty level of the learning program as appropriate, confirms that the quality of the adaptation and the estimation of the children's knowledge were sufficient. Moreover, the need of adaptation to the level of each child is demonstrated by the case studies (section Case studies). As seen in the pre-tests, each child starts with a different amount of mathematical knowledge and shows deficits in different areas. This is also reflected in the course of training: the path through the skill net varies across children. The case studies illustrate that children practice in areas, where they have deficits and generally demonstrate large improvements in these areas after training. Furthermore, it has been shown, that the use of a skill net allowing for different training trajectories optimizes the learning process (Käser et al., 2012).

The evaluation of the feedback questionnaire also supports the improvement of mathematical performance measured in the external tests: On average, children reported that the training had improved their mathematical performance. This subjective feeling of improvement and learning success might also enhance positive self-concepts (Ashcraft and Faust, 1994; Spitzer, 2009). Moreover, children also indicated that they liked to train with the program. The popularity of the learning environment is beneficial as training can only be successful if the children are motivated. Furthermore, the finding demonstrates that the computer is an attractive medium for children and is in line with previous studies (Kulik and Kulik, 1991; Schoppek and Tulis, 2010; Kucian et al., 2011).

BEHAVIORAL EFFECTS

Our first results reveal positive training effects in mathematical skills after completion of the training. Children significantly improved their subtraction skills over the course of the 6-weeks-training: They were not only able to solve more complex subtraction problems (medium-large effect in the computer-based subtraction test) but also solved subtraction tasks faster (large effect in HRT). This improvement in subtraction supports the notion of a better mathematical understanding as subtraction is considered as a main indicator for the development of the spatial number line representation (Dehaene, 2011). Furthermore, the decrease in problem solution times can be seen as a shift to increased fact retrieval (Geary et al., 1991; Lemaire and Siegler, 1995; Barrouillet and Fayol, 1998; Jordan et al., 2003). Compared to subtraction, children demonstrated smaller effects in addition (medium effect in computer-based addition test). This may be due to the adaptive nature of the intervention: Addition and subtraction tasks are trained in parallel for each difficulty level. As children performed better in addition at the pre-test, they received more training in subtraction during the intervention. Interestingly, the waiting group did not show significant training effects in the HRT subtests after their 6-weeks training ($t_2 - t_3$). This fact might stem from the low number of participants or from the adaptability of the training program leading to a different training trajectory for each child.

Children were also able to locate the position of a number on a number line more accurately after training. In the number range between 0 and 10, the deviation from the correct position was reduced by 33% after 6 weeks. Children especially also reduced the variance (medium-large effect size). No further improvement was yield by the prolongation of the training. Yet, most children passed the skills in the number range from 0 to 10 in the first few weeks and thus did not train in this range anymore in the second part of the training. In the number range between 0 and 100, there was no significant interaction. However, the training effect was significant after 3 months (reduction of deviation about 30%). This delay is probably due to the fact, that some of the children arrived at this level only in the second part of the training. Better performance in the number line task indicates refinement of the internal mental number line and more accurate access to it and confirms the results of previous studies (Siegler and Booth, 2004; Booth and Siegler, 2006, 2008; Halberda et al., 2008) which demonstrated significant correlations between arithmetical learning and the quality of numerical magnitude representation.

No significant training effects were observed in the NC and estimation tasks. These results however need to be interpreted with caution because of ceiling effects. At the pre-tests, children solved on average 80% of the NC and 75% of the estimation tasks correctly. Furthermore, some children even reached the maximum score. This result is in line with previous findings (Noel and Rousselle, 2011).

For most of the tasks tested before and after training, prolongation of the training from 6 to 12 weeks yielded a beneficial effect. The improvement of the training group over the whole training period ($t_1 - t_3$) was significant for all, but the estimation and NC tests. In some tasks (for example NL100 and NL1000), the effects of the second part of the training were similar or higher to those of the first part. This may be due to two facts. Firstly, as the training covers the number range from 0 to 1000, most children had not worked through the whole training after the first 6 weeks. Secondly, the intervention trains different abilities whose effects support each other. However, the supporting effects between those abilities need time to develop (Kaufmann et al., 2003, 2005). The prolongation of the training time from 6 to 12 weeks thus probably led to a strengthening of the mutual effects between the training in number representations and the training in arithmetic operations.

Although a training program focusing on a broad range of mathematical skills and showing a high degree of individualization seems beneficial, it also poses challenges for the evaluation. Firstly, training a variety of skills shortens the training time of each specific skill and thus leads to smaller training effects as mentioned above. Secondly, due to the high adaptability of the program, each child pursues a different training trajectory, i.e., the children train different skills and might even not train a specific skill at all because they either already possessed that ability prior to the training or did not arrive at this difficulty level during training. Therefore, training progress is hard to compare and inconsistencies in training effects may be observed. Nevertheless, the first obtained pilot results are promising and form the basis for further evaluation.

Another important point is the connection to practice. When used in school settings, computer-based training programs might not show significant improvement compared to conventional classroom teaching (Dynarski et al., 2007). However, we believe that carefully designed computer-based training programs provide a valuable addition to conventional classroom teaching in providing a possibility to differentially address individual characteristics.

LIMITATIONS

Some limitations regarding the participants and the study design have to be considered. Firstly, there were no measurements done after a 12-weeks rest period. Thus, for the 12-weeks training period, the training effects could not be compared to the effects of a rest period. Regarding the significant effects of the 6-weeks training, we conclude that also the effects of the 12-weeks training period can be plausibly attributed to the training.

Secondly, children were not tested according to common criteria of DD. Children were indicated by parents and teachers as exhibiting difficulties in learning mathematics. Generally, participants indeed demonstrated a mathematical performance below the 25th percentile in the pre-tests (the four children performing above the 25th percentile had insufficient grades in math). However, as described in section Study design and participants, the participants' mean score even demonstrated an arithmetic performance around the 10th percentile. A further study restricted to children diagnosed with DD is currently conducted in Germany. Nevertheless, our less strict criterion for deficits in mathematical performance seems also informative. It has been shown that the cognitive characteristics of low performing children are indeed dependent on the cut-off criterion used. However, children fulfilling a softer criterion exhibit similar difficulties to those fulfilling stronger criteria, but to a smaller extent (Murphy et al., 2007).

Thirdly, the effects of the training period were only compared to those of a rest period. No comparison to a control training was conducted. As this first pilot study was designed to evaluate the concepts used in the training program and to assess its adaptability, the design used seems sufficient. Having proved the effects of the training in this first step, the mentioned further study conducted in Germany will compare the effects also to a control training.

OUTLOOK

The presented results from the first evaluation form the basis for further evaluation and improvement of the training program. In a first step, the program is evaluated in a large-scale study conducted in Germany. This study compares effects to a control training and also assesses domain-unspecific measures such as attention, overcoming the limitations of the pilot study.

In a second step, the training program will be improved and extended.

Evaluation of input data and observations of supervised training sessions have shown that children advance too fast in the area of arithmetic operations. Therefore, an incorporation of answer times into the mathematical model is planned, allowing to set time limits for tasks and giving an indication of strategies used

(for example fact retrieval versus counting). Furthermore, games training different calculation strategies would be beneficial and put more emphasis on conceptual knowledge (instead of fact knowledge).

The current concept of the training program balances the training time between the area of arithmetic operations and number representations. However, while the area of arithmetic operations trains only addition and subtraction skills, a variety of skills are trained in the area of number representations. Due to this high number of skills, some skills are only trained for a short amount of time and thus no significant improvement in these skills can be observed. The external tests do for example not show any significant improvement in estimation or non-symbolic magnitude comparison tasks. Allocating more training time to the number representations area might solve this issue.

At the moment, the training program entirely relies on intrinsic motivation. Although children indicated that they liked to train with the program, the training could benefit from additional motivational instruments such as the collection of points (for example for correct tasks or training time) and the visualization of learning progress. A version including additional instruments is already planned and will be evaluated in a further user study.

CONCLUSION

In the present study, the computer-based training program Calcularis for children with mathematical learning problems was developed and evaluated. The design of the program is based on current neuropsychological findings. The program features a control algorithm allowing adaptation to the user and thus optimization of learning processes. Evaluation of the learning program showed significant training effects in number representation as well as in subtraction. The program proved to adapt well to the needs of the children and feedback from participants was positive. The results obtained from the first evaluation form a promising basis for further evaluation.

REFERENCES

Ansari, D., and Dhital, B. (2006). Age-related changes in the activation of the intraparietal sulcus during non-symbolic magnitude processing: an event-related functional magnetic resonance imaging study. *J. Cogn. Neurosci.* 18, 1820–1828. doi: 10.1162/jocn.2006.18.11.1820

Arnold, D. H., Fisher, P. H., Doctoroff, G. L., and Dobbs, J. (2002). Accelerating math development in head start classrooms. *J. Educ. Psychol.* 94, 762–770. doi: 10.1037/0022-0663.94.4.762

Arsalidou, M., and Taylor, M. J. (2011). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. *Neuroimage* 54, 2382–2393. doi: 10.1016/j.neuroimage.2010.10.009

Ashcraft, M. H., and Faust, M. W. (1994). Mathematics anxiety and mental arithmetic performance: an exploratory investigation. *Cogn. Emotion* 8, 97–125. doi: 10.1080/02699939408408931

Barrouillet, P., and Fayol, M. (1998). From algorithmic computing to direct retrieval: evidence from number and alphabetic arithmetic in children and adults. *Mem. Cogn.* 26, 355–368. doi: 10.3758/BF03201146

Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 42, 189–201. doi: 10.1037/0012-1649.41.6.189

Booth, J. L., and Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Dev.* 79, 1016–1031. doi: 10.1111/j.1467-8624.2008.01173.x

Butterworth, B., Varma, S., and Laurillard, D. (2011). Dyscalculia: from brain to education. *Science* 332, 1024.

Callaway, E. (2013). Dyscalculia: number games. *Nature* 493, 150–153. doi: 10.1038/493150a

Carey, S. (2001). Cognitive foundations of arithmetic: evolution and ontogenesis. *Mind Lang.* 16, 35–55. doi: 10.1111/1468-0017.00155

Carey, S. (2004). Bootstrapping and the origin of concepts. *Daedalus* 133, 59–68. doi: 10.1162/001152604772746701

Cattell, R. B., Weiss, R. H., and Osterland, J. (1997). *Grundintelligenztest Skala 1 (CFT 1).* Goettingen: Hogrefe Verlag.

Cohen Kadosh, R., Cohen Kadosh, K., Kaas, A., Henik, A., and Goebel, R. (2007). Notation-dependent and -independent representations of numbers in the parietal lobes. *Neuron* 53, 307–314. doi: 10.1016/j.neuron.2006.12.025

Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics.* New York, NY: Oxford University Press.

Dehaene, S., and Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Math. Cogn.* 1, 82–120.

Dowker, A. D. (2001). Numeracy recovery: a pilot scheme for early intervention with young children with numeracy difficulties. *Support Learn.* 16, 6–10. doi: 10.1111/1467-9604.00178

Dowker, A. D. (2003). "Interventions in numeracy," in *New Directions in Numeracy Education,* ed I. Thompson (Buckingham: Open University Press), 127–135.

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., et al. (2007). *Effectiveness of Reading and Mathematics Software Products: Findings from the First Student Cohort.* Washington, DC: U.S. Department of Education.

Field, A. (2009). *Discovering Statistics Using SPSS.* London: Sage.

Fletcher-Flinn, C. M., and Gravatt, B. (1995). The efficacy of computer assisted instruction (CAI): a meta-analysis. *J. Educ. Comput. Res.* 12, 219–241. doi: 10.2190/51D4-F6L3-JQHU-9M31

Fuchs, L. S., Fuchs, D., Hamlet, C. L., Powell, S. R., Capizzi, A. M., and Seethaler, P. M. (2006). The effects of computer-assisted instruction on number combination skill in at-risk first graders. *J. Learn. Disabil.* 39, 467–475. doi: 10.1177/00222194060390050701

Geary, D. C. (2004). Mathematics and learning disabilities. *J. Learn. Disabil.* 37, 4–15. doi: 10.1177/00222194040370010201

Geary, D. C., Brown, S. C., and Samaranayake, V. A. (1991). Cognitive addition: a short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Dev. Psychol.* 27, 787–797. doi: 10.1037/0012-1649.27.5.787

Gerster, H. D. (1982). *Schülerfehler bei schriftlichen Rechenverfahren - Diagnose und Therapie.* Freiburg: Herder.

Griffin, S. A., Case, R., and Siegler, R. S. (1994). "Rightstart: providing the central conceptual prerequisites for first formal learning of arithmetic to students at risk for school failure," in *Classroom Lessons: Integrating Cognitive Theory and Classroom Practice,* ed K. Mcgilly (Cambridge, MA: The MIT Press), 25–49.

Haffner, J., Baro, K., Parzer, P., and Resch, F. (2005). *Heidelberger Rechentest (HRT): Erfassung mathematischer Basiskompetenzen im Grundschulalter.* Goettingen: Hogrefe Verlag.

Halberda, J., Mazzocco, M. M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* 455, 665–668. doi: 10.1038/nature07246

Hannula, M. M., and Lehtinen, E. (2005). Spontaneous focusing on numerosity and mathematical skills of young children. *Learn. Instruct.* 15, 237–256. doi: 10.1016/j.learninstruc.2005.04.005

Jordan, N. C., Hanich, L., and Uberti, H. Z. (2003). "Mathematical thinking and learning disabilities," in *The Development of Arithmetic Concepts and Skills. Recent Research and Theory,* eds. A. Baroody and A. Dowker (Mahwah, NJ: Erlbaum), 359–389.

Karmiloff-Smith, A. (1992). *Beyond Modularity.* Cambridge, MA: MIT Press.

Käser, T., Busetto, A. G., Baschera, G.-M., Kohn, J., Kucian, K., Aster, M., et al. (2012). Modelling and optimizing the process of learning mathematics. *Proc. ITS* 7315, 389–398.

Kaufmann, L., Delazer, M., Pohl, R., Semenza, C., and Dowker, A. (2005). Effects of a specific numeracy educational program in preschool children: a pilot study.

Educ. Res. Eval. 11, 405–431. doi: 10.1080/1380361050011049

Kaufmann, L., Handl, P., and Thöny, B. (2003). Evaluation of a numeracy intervention program focusing on basic numerical knowledge and conceptual knowledge: a pilot study. *J. Learn. Disabil.* 35, 564–573. doi: 10.1177/00222194030360060701

Kaufmann, L., and von Aster, M. G. (2012). The diagnosis and management of dyscalculia. *Dtsch. Arztebl. Int.* 109, 767–778.

Kaufmann, L., Wood, G., Rubinsten, O., and Henik, A. (2011). Meta-analyses of developmental fMRI studies investigating typical and atypical trajectories of number processing and calculation. *Dev. Neuropsychol.* 36, 763–787. doi: 10.1080/87565641.2010.549884

Khalili, A., and Shashaani, L. (1994). The effectiveness of computer applications: a meta-analysis. *J. Res. Comput. Educ.* 27, 48–61.

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

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., et al. (2011). Mental number line training in children with developmental dyscalculia. *Neuroimage* 57, 782–795. doi: 10.1016/j.neuroimage.2011.01.070

Kucian, K., and Kaufmann, L. (2009). A developmental model of number representation. *Behav. Brain Sci.* 32, 313–373. doi: 10.1017/S0140525X09990069

Kucian, K., von Aster, M., Loenneker, T., Dietrich, T., and Martin, E. (2008). Development of neural networks for exact and approximate calculation: a fMRI study. *Dev. Neuropsychol.* 33, 447–473. doi: 10.1080/87565640801014474

Kulik, C.-L. C., and Kulik, J. A. (1991). Effectiveness of computer-based instruction: an updated analysis. *Comput. Hum. Behav.* 7, 75–94. doi: 10.1016/0747-5632(91)90030-5

Kulik, J. A. (1994). "Meta-analytic studies of findings on computer-based instruction," in *Technology Assessment in Education and Training*, eds E. L. Baker and H. F. O'Neil (Hillsdale, NJ: LEA Publishers), 9–33.

Kullik, U. (2004). "Computer-unterstützte rechen trainings programme," in *Interventionen bei Lernstörungen*, eds G. W. Lauth, M. Grünke, and J. C. Brunstein (Göttingen: Hogrefe), 329–337.

Landerl, K., Bevan, A., and Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8-9-year-old students. *Cognition* 93, 99–125. doi: 10.1016/j.cognition.2003.11.004

Lemaire, P., and Siegler, R. S. (1995). Four aspects of strategic change: contributions to children's learning of multiplication. *J. Exp. Psychol. Gen.* 124, 83–97. doi: 10.1037/0096-3445.124.1.83

Lenhard, A., Lenhard, W., Schug, M., and Kowalski, A. (2011). Computerbasierte Mathematikförderung mit den "Rechenspielen mit Elfe und Mathis I". *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 43, 79–88. doi: 10.1026/0049-8637/a000037

Li, Q., and Ma, X. (2010). A meta-analysis of the effects of computer technology on school students' mathematics learning. *Educ. Psychol. Rev.* 22, 215–243. doi: 10.1007/s10648-010-9125-8

Mazzocco, M. M., and Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learn. Disabil. Res. Pract.* 20, 142–155. doi: 10.1111/j.1540-5826.2005.00129.x

Murphy, M. M., Mazzocco, M. M., Hanich, L., and Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cut-off criterion used to define MLD. *J. Learn. Disabil.* 40, 458–478. doi: 10.1177/00222194070400050901

Naglieri, J. A., and Johnson, D. (2000). Effectiveness of a cognitive strategy intervention in improving arithmetic computation based on the PASS theory. *J. Learn. Disabil.* 33, 591–597. doi: 10.1177/002221940003300607

Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proc. Natl. Acad. Sci. U.S.A.* 109, 6289–6294.

Noel, M.-P., and Rousselle, L. (2011). Developmental changes in the profiles of dyscalculia: an explanation based on a double exact-and-approximate number representation model. *Front. Hum. Neurosci.* 5:165. doi: 10.3389/fnhum.2011.00165

Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., and Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cognit. Dev.* 24, 450–472. doi: 10.1016/j.cogdev.2009.09.003

Rivera, S. M., Reis, A. L., Eckert, M. A., and Menon, V. (2005). Developmental changes in mental arithmetic: evidence for increased functional specialization in the left interior parietal cortex. *Cereb. Cortex* 15, 1779–1790. doi: 10.1093/cercor/bhi055

Rubinsten, O., and Sury, D. (2011). Processing ordinality and quantity: the case of developmental dyscalculia. *PLoS ONE* 6:e24079. doi: 10.1371/journal.pone.0024079

Schoppek, W., and Tulis, M. (2010). Enhancing arithmetic and word-problem solving skills efficiently by individualized computer-assisted practice. *J. Educ. Res.* 103, 239–252. doi: 10.1080/00220670903382962

Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi: 10.1111/j.1467-8624.2004.00684.x

Spitzer, M. (2009). Ja, ich kann! Selbstbild, Selbstbejahung und nachhaltige Leistungsfähigkeit. *Nervenheilkunde* 28, 425–430.

Van De Rijt, B. A. M., and Van Luit, J. E. H. (1998). Effectiveness of the additional early mathematics program for teaching children early mathematics. *Instr. Sci.* 26, 337–358. doi: 10.1023/A:1003180411209

Van Luit, J. E. H., and Naglieri, J. A. (1999). Effectiveness of the MASTER program for teaching special children multiplication and division. *J. Learn. Disabil.* 32, 98–107. doi: 10.1177/002221949903200201

Vogel, S. E., and Ansari, D. (2012). Neurokognitive Grundlagen der typischen und atypischen Zahlenverarbeitung. *Lernen und Lernstörungen* 1, 135–149.

von Aster, M. G. (2000). Developmental cognitive neuropsychology of number processing and calculation: varieties of developmental dyscalculia. *Eur. Child Adolesc. Psychiatry* 9, S41–S57.

von Aster, M. G., and Shalev, R. S. (2007). Number development and developmental dyscalculia. *Dev. Med. Child Neurol.* 49, 868–873. doi: 10.1111/j.1469-8749.2007.00868.x

Weiss, R. H. (2006). *Grundintelligenztest Skala 2 - Revision - (CFT 20-R)*. Goettingen: Hogrefe Verlag.

Wilson, A. J., and Dehaene, S. (2007). "Number sense and developmental dyscalculia," in *Human Behavior, Learning, and the Developing Brain: A typical Development*, eds D. Coch, G. Dawson, and K. Fischer (New York, NY: Guilford Press), 212–238.

Wilson, A. J., Dehaene, S., Dubois, O., and Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind Brain Educ.* 3, 224–234. doi: 10.1111/j.1751-228X.2009.01075.x

Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S. K., Cohen, L., and Cohen, D. (2006a). Principles underlying the design of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behav. Brain Funct.* 2:19. doi: 10.1186/1744-9081-2-19

Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., and Dehaene, S. (2006b). An open trial assessment of "The Number race", an adaptive computer game for remediation of dyscalculia. *Behav. Brain Funct.* 2:20. doi: 10.1186/1744-9081-2-20

Wright, R. J. (2003). A mathematics recovery: program of intervention in early number learning. *Aust. J. Learn. Disabil.* 8, 6–11. doi: 10.1080/19404150309546741

Examining the presence and determinants of operational momentum in childhood

André Knops[1], Steffen Zitzmann[1] and Koleen McCrink[2]*

[1] *Department of Psychology, Humboldt University at Berlin, Berlin, Germany*
[2] *Barnard College, Columbia University, New York, NY, USA*

The operational momentum (OM) effect describes a systematic bias in estimating the outcomes of simple addition and subtraction problems. Outcomes of addition problems are overestimated while outcomes of subtraction problems are underestimated. The origin of OM remains debated. First, a flawed uncompression of numerical information during the course of mental arithmetic is supposed to cause OM due to linear arithmetic operations on a compressed magnitude code. Second, attentional shifts along the mental number line are thought to cause OM. A third hypothesis explains OM in 9-month olds by a cognitive heuristic of accepting more (less) than the original operand in addition (subtraction) problems. The current study attempts to disentangle these alternatives and systematically examines potential determinants of OM, such as reading fluency which has been found to modulate numerical–spatial associations. A group of 32 6- and 7-year-old children was tested in non-symbolic addition and subtraction problems, in which they had to choose the correct outcome from an array of several possible outcomes. Reading capacity was assessed for half of the children while attentional measures were assessed in the other half. Thirty-two adults were tested with the identical paradigm to validate its potential of revealing OM. Children (and adults) were readily able to solve the problems. We replicated previous findings of OM in the adults group. Using a Bayesian framework we observed an inverse OM effect in children, i.e., larger overestimations for subtraction compared to addition. A significant correlation between children's level of attentional control and their propensity to exhibit OM was observed. The observed pattern of results, in particular the inverse OM in children is hard to reconcile with the previously proposed theoretical frameworks. The observed link between OM and the attentional system might be interpreted as evidence partially supporting the attentional shift hypothesis.

Keywords: approximate calculation, non-symbolic calculation, mental number line, development, space and numbers, attention, numerical cognition

INTRODUCTION

Along with a variety of species humans possess an untrained and non-symbolic "number sense," which yields representations of numerical magnitude that can then be used productively in arithmetic operations such as addition and subtraction (Gallistel, 1990; Wynn, 1992; McCrink and Wynn, 2004; Barth et al., 2005; Cordes et al., 2007; Nieder and Dehaene, 2009). When estimating the outcome of simple mental calculation problems, adults systematically overestimated the outcomes of addition problems and underestimated the outcomes of subtraction problems (McCrink et al., 2007). Analogous to a perceptual phenomenon called representational momentum (Freyd and Finke, 1984) – in which adults misperceive the position at which a moving object disappears in the direction of the movement – this effect was termed operational momentum (OM). OM was observed for both symbolic (e.g., Arabic numerals) and non-symbolic notations (e.g., arrays of objects), implying a notation-independent mechanism which uses semantic and abstract magnitudes as input (Knops et al., 2009b). OM is also observed in paradigms using different response modalities (i.e., choosing from a number of responses, or pointing to the estimated outcome on a linear number scale), suggesting that its origin lies at a central cognitive processing level (Pinhas and Fischer, 2008).

In order to properly detail the existing theoretical hypotheses which attempt to explain OM in approximate calculation, we must first introduce two notions that are crucial for the understanding of the proposed mechanisms. Originally (McCrink et al., 2007; Knops et al., 2009b) the OM bias was explained by mechanisms which describe the underlying numerical magnitude representations as: (a) logarithmically compressed and (b) spatially oriented, with smaller numbers located left from larger numbers. The adult humans tested in these studies possess a cognitive system that enables them to perceive and process numerical magnitude information in an approximate, analog fashion – the aforementioned "number sense," or approximate number system (ANS). The ANS yields a sense of a given numerical magnitude by activating a fixed position along a numerically ordered continuum, commonly referred to as the Mental Number Line (MNL). Crucially, the MNL

is hypothesized to be logarithmically compressed; that is, distances between neighboring numbers decrease logarithmically inversely to the numbers' magnitudes. Due to the noisiness of activation signals in the ANS, activation at a given position on the MNL will also partially activate adjacent positions. Although still under debate, mounting evidence from behavioral (Moyer and Landauer, 1967; Izard and Dehaene, 2008) and computational studies (Dehaene and Changeux, 1993) suggests that the mental magnitude representation is logarithmically compressed. Most central to the current study, recent data from single-unit recordings has been used to directly test the assumption of a logarithmic compression with fixed variability against alternative scales such as a linear scale with increasing variability (Nieder and Miller, 2003). Models assuming compressed scales yielded better fit indices than a linear scale both during perception of number stimuli and during maintenance in memory. Moreover, compressed scaling of symbolic numbers has been demonstrated to persist in educated Western adults (Viarouge et al., 2010). In a series of experiments adults were asked to judge whether a given sequence of numbers contained too many small numbers or too many large ones. Participants judged as random those sequences that oversampled small numbers. And finally, while scaling of symbolic numbers may linearize over time (Siegler and Opfer, 2003), possibly due to education, non-symbolic numerosities have been found to be mapped in a non-linear way in adults (Dehaene et al., 2008). Nevertheless, some researchers suggest a linearly scaled mental magnitude representation with increasing variability as numerical magnitude increases (Gallistel and Gelman, 2000; Brannon et al., 2001; Ebersbach et al., 2008; Gallistel, 2011; Stoianov and Zorzi, 2012). Indeed, when asking children to place numbers on a spatial scale (e.g., a line) according to their cardinal value the observed mappings change from a logarithmic mapping scheme to a linear mapping scheme as a function of number knowledge and familiarity with numerical concepts (Siegler and Opfer, 2003). Children exhibited a linear mapping scheme for familiar number ranges (e.g., 1–100 for second and fourth graders) and a logarithmic mapping scheme in an unfamiliar number range (e.g., 1–1000 for second and fourth graders) (Siegler and Opfer, 2003; Berteletti et al., 2010). It is unclear to what extent these mapping schemes reflect the scaling schema of the underlying representation, however (Karolis et al., 2011). In sum, we think there is good evidence for a compressed numerical magnitude representation with fixed variability, especially for non-symbolic numerosity information.

Several lines of evidence support the notion of a spatially oriented mental magnitude representation. The classic Spatial–Numerical Association of Response Codes (SNARC) effect implies an association of numerical magnitude representation with external space (Dehaene et al., 1993). In this phenomenon, left-side responses are faster for small numbers, while right-side responses are faster for larger numbers, providing evidence for a spatially oriented MNL in which adults associate small numbers with the left side of space and large numbers with the right. The left-to-right orientation was suggested to result from reading habits in particular societies, such as the French-speaking (left-to-right reading and writing) sample tested by Dehaene et al. (1993); the phenomenon was attenuated in Iranian participants (right-to-left reading and writing) in relation to the number of years they had been in

France (Dehaene et al., 1993). Shaki et al. (2009, p. 331) found that the SNARC effect was reversed in Palestinian participants who read both words and numbers from right to left, bolstering the idea that directional reading habits (e.g., left-to-right in Western cultures) "enables the association between numbers and space to become significant," which in turn may lead to differentially oriented mental number representations depending on cultural and situational variables (Bächtold et al., 1998). The cultural impact on the SNARC effect and the assumed underlying MNL representation is also supported by studies on the developmental trajectory of this phenomenon. The majority of studies investigating the spontaneous spatial orientation of the MNL in children using classic SNARC-like tasks failed to observe significant results before the age of 9 years (Berch et al., 1999; van Galen and Reitsma, 2008; Imbo et al., 2012). However, some recent work using child-friendly paradigms has called this into question, with some evidence for culturally appropriate spatial mapping of small-large magnitudes as early as the preschool years (Opfer et al., 2010; Patro and Haman, 2012). Recent evidence suggests that initial spatial biases become strengthened or weakened depending on the nature of the schooling that children receive (Shaki et al., 2012). Thus, it is conceivable that existing spatial biases consolidate with increasing reading proficiency. Further, semantic activations of magnitudes cause spatial shifts of attention. Fischer et al. (2003) found that the numerical magnitude of numbers presented centrally before a simple stimulus detection task had a systematic impact on participants' performance. Participants responded faster to left-sided targets than to right-sided targets when targets followed the presentation of small numbers. An equivalent advantage was observed for right-sided stimuli following large numbers. Similarly, Nicholls et al. (2008) demonstrated that participants were biased in their decision about which of two gray scales was darker by the numerical magnitude of superimposed digits. Left- and right-ward attentional biases were observed for low and high numbers, respectively. In the line-bisection effect, participants are relatively accurate at finding the midpoint of a line comprised of a series of "x"s, but deviate left- or right-ward when the line is comprised of a string of the word "two" or "nine," respectively (Fischer, 2001; Calabria and Rossetti, 2005). Finally, damage to right parietal cortex elicits visuo-spatial hemineglect alongside representational neglect of portions of the MNL (Zorzi et al., 2002). Patients suffering from hemineglect not only misperceived visual information from the left hemifield, they also neglected numerical information from the left side of the MNL; when asked for the numerical middle between 1 and 9 the patients responded "6," as if they did not consider the smaller numbers located on the left side of the mental number representation.

Together, the above findings strongly support the notion of a left-to-right oriented and logarithmically compressed MNL. This construct is central when considering the nature of the mechanisms put forth to explain the phenomenon of OM. Here we will detail three such proposed mechanisms, which are not necessarily exclusive of each other. The first mechanism was proposed by McCrink et al. (2007) in their original documentation of OM, and implemented in a computational model by Chen and Verguts (2012). It is based on two notions: first, it assumes a compressed mental magnitude representation. Second,

it assumes that the cognitive system "undoes" the compression during mental calculation and operates on uncompressed magnitudes. This process of uncompression may be subject to a systematic bias which results in a slightly compressed magnitude code during calculation. This compressive bias may in turn cause the OM. A simple example illustrates this idea. Imagine a participant adds two numbers, e.g., $20 + 5$. Internally, these are represented as $\log 10(20) = 1.301$ and $\log 10(5) = 0.699$. In the most extreme case, the uncompression process fails completely and participants will actually operate on the log-scaled values and add $\log 10(20) = 1.301$ and $\log 10(5) = 0.699$. Adding two logarithms corresponds to multiplying their linear-scaled values, i.e., $\log 10(20 + 5) \sim 20 \times 5 = 100$ and in most cases this would result in values larger than the actual outcome. A similar argument holds for subtraction which would be replaced by division. Note that this example is used only to illustrate the basic idea of this account. The actually observed biases are much smaller than in this example. The main idea is that participants apply a linear transformation on a compressed scale which will lead to over- or under-estimating the outcome of a given problem. This hypothesis will be referred to as the "compression account."

The second account appeals to attentional shifts along the MNL. According to this hypothesis, arithmetic operations are mediated by a dynamic interplay between cortical structures which process spatial information. In particular, it has been reasoned that a bilateral circuitry involving posterior superior parietal lobe (PSPL) and horizontal intraparietal sulcus (hIPS) that implements a form of vector addition over eye and retinal position information is co-opted by mental arithmetic. Indeed, exploiting the fact that saccades are accompanied by shifts of spatial attention in saccade direction, Knops et al. (2009a) used the brain activity elicited by left- and right-ward saccades to predict whether French-speaking participants were performing centrally presented addition or subtraction problems. The authors found that addition problems corresponded to the neural activity associated with right-ward saccades, presumably since participants shift attention toward larger numbers on the right side of the MNL. OM results from the momentum that drives participants too far along the MNL in the direction of the operation. This hypothesis will be referred to as the "attentional shifts account."

Finally, a third hypothesis was proposed by McCrink and Wynn (2009) to explain the possible presence of OM in a population of 9-month-old infants. The authors presented visual sequences of addition and subtraction problems using non-symbolic numerosities (such as $6 + 4 = 5$, 10, or 20), and provided the infants with three different types of outcomes to these problems: correct, too large, or too small. The infants looked reliably longer to the outcomes that violated the "momentum" of the particular problem. Infants who saw an addition scenario looked for a relatively long time at outcomes that were too small, but similarly to outcomes that were correct and too big; infants who saw a subtraction scenario looked longer at outcomes that were too large, and less to the correct and too-small outcomes. The authors put forward two suggestions to explain this pattern, the first was a computational account in which the infants are genuinely computing

an acceptable outcome with some amount of error in magnitude representations that went in the "direction" of the operation. However, since a full-fledged, spatially oriented MNL is unlikely in this relatively unencultured population, the authors hypothesized that they may instead be deploying general arithmetic principles. Specifically, "if adding, accept more" than the original operand, and "if subtracting, accept less (McCrink and Wynn, 2009, p. 407)." We will refer to this hypothesis as the "heuristics account."

No study to date has attempted to disentangle these alternatives. In the following experiment we systematically examine potential determinants of OM, looking at a population that serves as a transition group between infants and adults-children in their first year of school. Six- and seven-year-old children were given a series of non-symbolic addition and subtraction problems, in which they had to choose the correct outcome from an array of several possible outcomes. The children were placed in either a reading condition or a cueing condition, in which we also separately tested their reading automaticity or attentional orienting capacity, respectively. (A group of adults was also tested on the non-symbolic addition and subtraction task, to ensure the efficacy of the paradigm in eliciting OM.) This design allows us to address two outstanding questions in the literature. First, do we see any evidence for a MNL at this age, as exhibited by the overall presence of a spatial–numerical interaction during arithmetic operations (OM)? Second, insofar as they exhibit OM, what are the determinants of the presence of this phenomenon?

Our predictions are as follows. First, if the MNL is instantiated via the highly automatic and culturally directed reading habits of the children, we will see a positive relationship between reading fluency and level of exhibited OM. Reading has been shown to modulate number-space associations while illiterate individuals did not exhibit consistent associations between numbers and space (Shaki et al., 2012). If there is not a spatially organized MNL at this age, regardless of reading ability, or if the MNL is present but not dictated by the child's literacy, we will observe no relationship between these two constructs. Second, if OM is due to the flawed uncompression of mental numerosities during the course of mental arithmetic, we should observe a standard adult-like OM effect in 6- and 7-year olds, irrespective of their measures on reading or attention. Ample evidence from line-bisection tasks supports a logarithmically compressed magnitude representation early in formal schooling, with only a prolonged shift to linear representations culminating in sixth grade – and then only for scales that have become familiar (Siegler and Opfer, 2003; Opfer and Siegler, 2007; Barth and Paladino, 2011). The distribution of responses should resemble the pattern shown by adults and peak at or around the correct outcome with a higher degree of acceptance of over- or under-estimated outcomes (for addition and subtraction, respectively). There will be a fall-off as the incorrect outcomes become extremely discrepant from the correct outcomes. Third, if the children's responses are driven by a "if adding, accept more, if subtracting, subtract less" heuristic account, we would expect to find an OM bias whose effect size would resemble the effect size found in adults. However, since children would not be engaging in an approximate calculation process, but rather show a general tendency to choose larger outcomes for addition and smaller

outcomes for subtraction, the distribution of responses would differ from what is observed in adults. The distribution would not necessarily be centered on or around the correct response in a given set of presented response alternatives. Rather, we should see a trend to accept as correct any amount larger than the first operand (for addition) or smaller than the first operand (for subtraction). Again, we would expect no effect of reading ability or attentional indices if this pattern was observed. Finally, if OM is underlain by shifts of spatial attention along a MNL, we will see a relationship between attentional indices and the strength and/or presence of OM. Children who exhibit strong orienting responses in the presence of spatial cues may have more adult-like circuits for deploying attention, and this results in adult-like OM in which *these particular children* peak in response choice relatively close to the correct answer, with a margin of acceptance for somewhat larger outcomes during addition problems and somewhat smaller outcomes for subtraction problems. At least two attentional functions can be distinguished that are relevant in the present context. Attentional selection refers to the observed benefit in performance in response to stimuli appearing at attended locations. When a stimulus appears at unattended locations, however, attention has to be re-oriented toward the formerly unattended location. This reorientation is time-consuming and hence responses to stimuli at unattended locations are delayed. Neuroimaging studies point to distinct cortical circuits for selection and reorienting (Corbetta et al., 2008).

MATERIALS AND METHODS
PARTICIPANTS
Children

The final sample consisted of 32 children (15 boys, 17 girls) between 6 ($n = 15$) and 7 ($n = 17$) years of age [mean (SD) age: 6.5 (0.51) years]. An additional four children were tested but excluded due to refusal to complete the study (3) or to homogenize the age range of the final sample (1). We also had an exclusion criterion for any children who were fluent in a right-to-left language (e.g., Hebrew); no one tested met this criteria. These children were recruited in the greater NYC area via word of mouth or through a local school. Informed written consent from the parents was obtained for all tested children. All children participated in a non-symbolic arithmetic task. Half of the sample (16) additionally participated in an attentional cueing paradigm [mean (SD) age: 6.38 (0.5) years], and the other half participated in a reading test [mean (SD) age: 6.69 (0.48) years]. In order to eliminate any interactions between the primary and secondary tasks, the order of the tasks was counterbalanced such that half the children received the primary adding/subtracting task first, and the other half received the secondary attention/reading task first.

Adult controls

A final sample of 32 college students (4 males, 28 females), recruited via the introductory psychology subject pool at a NYC university, were tested with the same non-symbolic calculation paradigm as the children. An additional eight students were excluded, seven for fluency in a right-to-left language (e.g., Hebrew), and one to keep sample size comparable between both groups. All students gave written informed consent and had

normal or corrected-to-normal sight. Mean age was 18.88 years (SD = 2.06).

TASKS
Arithmetic task
Stimuli. Four addition and four subtraction problems were created. Problems are presented in **Table 1**. To evaluate the differential effect of the arithmetic operation irrespective of numerical size of the outcome, we chose problems such that both arithmetic operations covered the same numerical range of final outcomes. Additionally, three memory trials were created with a second operand of zero to assess the overall capacity to retain the shown numerosities in mind.

Apart from the correct result, eight deviant results were created for each arithmetic problem. These deviants were arranged as a geometric series (i.e., were linearly spaced on a logarithmic scale) and ranged from double the correct result to half of the correct result [technically, they were generated as round ($c \times 2^i/4$), where c is the correct result and i ranges from -4 to $+4$]. Since this procedure would lead to some identical response alternatives for memory trials with set size = 6 (rounded deviants: [3, 4, 4, 5, 7, 8, 10, 12]) due to rounding to integers we repeatedly subtracted one from the smallest duplicate deviant until no further deviants were present. This resulted in the following response alternatives: [2, 3, 4, 5, 7, 8, 10, 12]. To avoid a strategy of always selecting the response falling in the middle of the proposed range, only six out of those nine possible results were presented on screen. In 50% of the trials we presented the upper six (high range), and thus the correct result was the second smallest numerosity (although numerosities were randomly mixed). In the other 50% of the trials the lower six choices were shown (low range), and the correct result was therefore the fifth smallest numerosity. For example, for a problem such as $6 + 2 = 8$ the low range would correspond to response alternatives 4, 5, 6, 7, 8, and 10. The high range would correspond to response alternatives 7, 8, 10, 11, 13, and 16.

Table 1 | Arithmetic problems presented in the non-symbolic calculation experiment and their correct and deviant results.

Operands		Results and deviant response alternatives								
First	Second	1/2	1/1.7	1/1.4	1/1.2	1	1.2	1.4	1.7	2
ADDITION										
6	2	4	5	6	7	8	10	11	13	16
6	4	5	6	7	8	10	12	14	17	20
14	5	9	11	13	16	19	23	27	32	38
14	11	13	15	18	21	25	30	35	42	50
SUBTRACTION										
16	8	4	5	6	7	8	10	11	13	16
16	6	5	6	7	8	10	12	14	17	20
32	13	9	11	13	16	19	23	27	32	38
32	7	13	15	18	21	25	30	35	42	50
MEMORIZATION										
6		2	3	4	5	6	7	8	10	12
19		9	11	13	16	19	23	27	32	38
25		13	15	18	21	25	30	35	42	50

To prevent the use of non-numerical cues, the sets of dots representing the non-symbolic numerosities were designed and generated using Matlab® such that dot size changed, but total dot area in a given set was always fixed across stimuli for half of the trials. As a result of this manipulation, average item size covaried inversely with numerosity during the presentation of the operands (i.e., sets with smaller numerosities had larger dots). For the second half of the trials individual dot size was held constant and total dot area covaried with numerosity (i.e., sets with smaller numerosities had smaller total dot area). Thus, neither total occupied area nor individual dot size could serve as a cue for distinguishing between the different numerosities throughout the experiment. To avoid memorization effects due to repetition of a particular stimulus, on each trial stimulus images were randomly chosen from a set of 10 precomputed images with the given numerosity.

PROCEDURE

The non-symbolic calculation task contained addition, subtraction, and memory trials. Each trial started with the presentation of a picture of a monkey that disappeared after a mouse button was clicked. An empty wooden box appeared at the bottom of the screen and the first set of dots moved into the box. The dot set appeared at the top of the screen and moved toward the middle of the screen with decreasing speed such that it would briefly remain stationary at the screen center before speed increased again and the dot disappeared inside the box. For addition problems a second set of dots appeared on screen and disappeared inside the box in the described manner. For subtraction problems a set of dots moved out of the box and disappeared at the top of screen. For memory trials only one set of dots disappeared in the box. After all sets of dots disappeared, three response alternatives appeared on the left and right side of the screen in an elliptic fashion, randomized for spatial location (i.e., the quantities are randomly assigned to the six positions). Each response alternative was presented as the top view of the box which contained different numbers of dots (see above). Children were asked to choose the correct outcome by clicking on the respective box. The beginning of the response active period was indicated by the appearance of the mouse pointer on top of a green star in the center of the screen. See **Figure 1** for a depiction of an exemplar addition trial. A total of 44 trials were presented, 12 memory trials, 16 addition, and 16 subtraction trials. A training period consisting of eight trials preceded the actual paradigm. In the training period, responses were not time-limited and feedback was provided. A correct choice was indicated by a green frame around the chosen box. The appearance of a red frame around the chosen box indicated that the choice was incorrect. During training response alternatives remained on screen until the correct response alternative was chosen. For the test period we limited response time to a maximum of 20 s. In the testing period, no feedback about the correctness of the choice was provided. The chosen alternative was highlighted by a surrounding blue frame, irrespective of correctness. On average children completed all calculation trials in about 20 min.

Attention task

Stimuli and procedure. An adapted Posner paradigm was administered to half of the child sample. A capital letter X in the center of the screen served as fixation point. On each trial the fixation was replaced by a yellow smiling face after a random interval between 1200 and 2400 ms. The smiling face indicated a forthcoming trial and remained visible for 750 ms. It was accompanied by an acoustic signal that served to attract children's attention. After a 250-ms interval with no stimuli on screen a blue arrow appeared in the center of the screen that pointed either to the left, to the right, or in both directions. The bidirectional arrow served as neutral condition. After a variable delay (200 or 800 ms) a blue star (introduced as "bug" to the children) appeared lateral to the arrow. The children were instructed to "zap the bug" by pushing the response button on the side of the target stimulus as fast as possible. The single headed arrows provided valid information about the side of the upcoming target in 67% ($n = 40$) of the trials (valid trials). In invalid trials ($n = 10$; 17%) the target stimulus appeared opposite the pointing direction of the arrow. The neutral condition was presented in 17% of the trials ($n = 10$). Target side (left, right) was balanced (50% each) in valid, invalid, and neutral trials. When the children hit the response button the target disappeared and an acoustic sound indicated whether the response was correct (sound 1) or not (sound 2). Eight training trials preceded the test and served to illustrate the task. The task lasted about 5–8 min.

This design allowed us to evaluate two central parameters of the attentional system: orienting (selection) and reorienting. Adopting common procedure from the attention literature the orienting effect was computed as the difference in reaction times between

FIGURE 1 | Screenshots taken during an addition trial. Two dot clouds sequentially move from the top of the screen into the box before the inside of the box is shown with six response alternatives. Children were asked to indicate the correct outcome by clicking on the respective box.

neutral and valid trials, the reorienting effect was computed as the difference in reaction times between neutral and invalid trials. All reaction time analyses were based on correct responses only.

Reading task

Stimuli and procedure. In order to assess reading ability and fluency, each child was provided with a developmentally appropriate vignette to read aloud while the experimenter tape-recorded him/her on the computer. The children were instructed to "Please read as much of this as you can, trying not to make any mistakes." The vignette was as follows:

> Mother and Father frog were going to a party. Mrs. Turtle came to babysit.
> "Hello, little frogs," said Mrs. Turtle, "What are we going to do tonight? Would you like me to read you a story?"
> "Yes! Yes!" said the little frogs, "we would like that very much."
> Mrs. Turtle finished reading. The little frogs cried, "Would you like to jump with us?"
> "Not now," said Mrs. Turtle, "It's suppertime. I will make you a nice supper."
> "Ok!" said the little frogs, "we are very hungry."

A coder blind to the hypotheses of the experiment reviewed the tapes of the children and assessed the subjective relative level of fluency for each child (from 1 to 10), and also coded how long each child took to read the vignette. Because we found that the time-to-read measure captured fluency in a more objective way than the coder rating, we used those data as each child's fluency score in our analyses.

DATA ANALYSIS

Data were analyzed using SPSS® under a classical statistical null hypothesis significance testing (NHST) approach. Bayesian analyses were conducted using R (R Development Core Team, 2010).

RESULTS
CHILDREN
Arithmetic task

Can children memorize the operands? First, we demonstrate that children were indeed capable to process and memorize the presented numerosities, albeit in an approximate fashion. If so, the mean chosen value on memory trials should closely follow the presented numerosities. **Figure 2** shows that this is actually the case. Mean chosen numerosity (squares) increased significantly with presented numerosity [$F(2, 60) = 279.65$, $p < 0.001$, epsilon $= 0.88$ (Huynh and Feldt, 1970)]. This main effect did not interact with age [$F(2, 60) = 1.17$, $p = 0.32$], indicating that both age groups were capable of remembering the presented numerosities. In line with the assumption of the mental magnitude representation following Weber's law we observed a constant coefficient of variation that did not significantly covary with numerosity [$F(2, 60) = 2.20$, $p = 0.12$; lower part of **Figure 2**]. No significant interaction with age was observed [$F(2, 60) < 1$]. A slight tendency to overestimate the remembered numerosities was present in the data. **Figure 2** (right) depicts the difference between mean chosen numerosity and the memorized numerosity. To test whether this difference was statistically different from zero we computed a
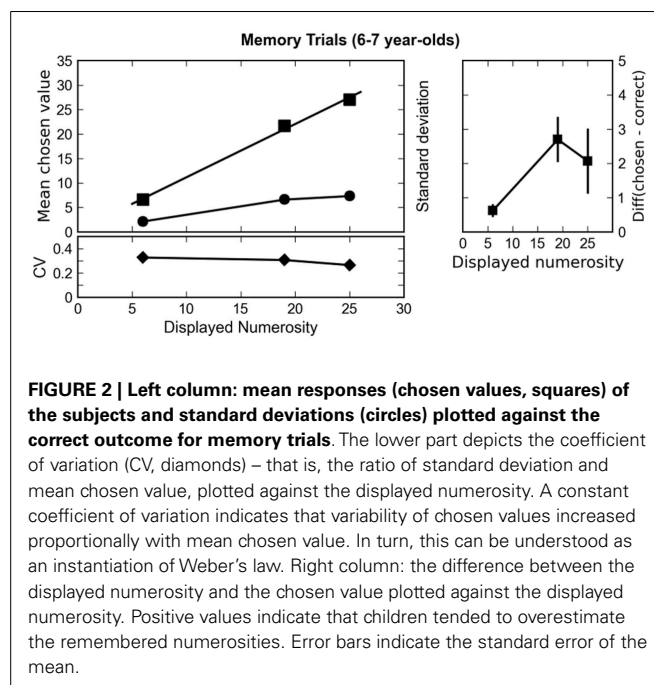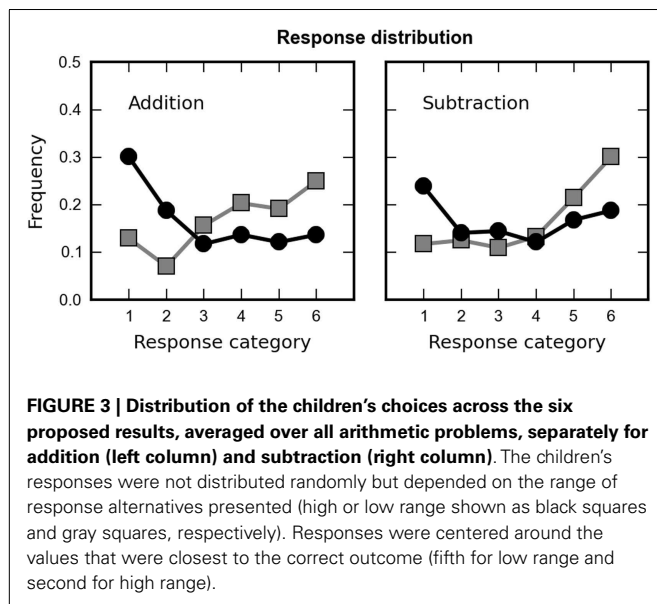

FIGURE 2 | Left column: mean responses (chosen values, squares) of the subjects and standard deviations (circles) plotted against the correct outcome for memory trials. The lower part depicts the coefficient of variation (CV, diamonds) – that is, the ratio of standard deviation and mean chosen value, plotted against the displayed numerosity. A constant coefficient of variation indicates that variability of chosen values increased proportionally with mean chosen value. In turn, this can be understood as an instantiation of Weber's law. Right column: the difference between the displayed numerosity and the chosen value plotted against the displayed numerosity. Positive values indicate that children tended to overestimate the remembered numerosities. Error bars indicate the standard error of the mean.

linear regression ($y = a + bx$) to predict memorized numerosity (y) based upon shown numerosity (x). If children were systematically overestimating the numerosities the intercept (a) of this regression equation would be significantly larger than zero. Mean intercept ($a = 0.63$) did not differ significantly from zero [$t(31) = 1.72$, $p = 0.095$]. However, mean slope ($b = 0.28$) was significantly larger than zero [$t(31) = 8.65$, $p < 0.001$], indicating that estimates increased with shown numerosity.

Did children engage in approximate calculation or respond at random? Next, we analyzed whether the subjects chose among the proposed choices at random. On each trial, six response alternatives were presented. They were either sampled from the lower range of response alternatives (alternatives one through six in **Table 1**) or the upper range of response alternatives (alternatives 3 through 9 in **Table 1**). As a consequence the correct outcome was either the second (high range) or the fifth smallest response alternative (low range) on screen. If the subjects were able to solve the arithmetic problems, their response choices should show a non-flat distribution, presumably centered close to the correct value. In contrast, if they responded randomly, we would not expect any differences in the frequency of choosing a particular response alternative. In **Figure 3**, we plot response frequency for each operation, separately for trials in which the correct answer was second (black) or fifth (gray). Responses were clearly distributed non-randomly. The peak of the distribution was always centered on response alternatives close to the correct outcome.

These conclusions were supported by an analysis of variance (ANOVA) over the different response categories, with (arcsine-transformed) percentage of choice as the dependent variable and rank of the subject's choice (one to six), range (second or fifth value correct), and operation (addition, subtraction) as factors. A main effect of rank [$F(5, 155) = 3.80$, $p = 0.012$, epsilon $= 0.62$]

FIGURE 3 | Distribution of the children's choices across the six proposed results, averaged over all arithmetic problems, separately for addition (left column) and subtraction (right column). The children's responses were not distributed randomly but depended on the range of response alternatives presented (high or low range shown as black squares and gray squares, respectively). Responses were centered around the values that were closest to the correct outcome (fifth for low range and second for high range).

FIGURE 4 | Left column: mean responses (chosen values, squares) of the children and standard deviations (circles) plotted against the correct outcome for addition (black) and subtraction problems (gray). The lower part depicts the coefficient of variation (CV, diamonds) – that is, the ratio of standard deviation and mean chosen value, plotted against the correct outcome. A constant coefficient of variation indicates that variability of chosen values increased proportionally with mean chosen value. In turn, this can be understood as an instantiation of Weber's law. Right column: the logarithm of the correct outcome plotted against the logarithm of the mean value chosen by the children for addition (black) and subtraction (gray). The gray line indicates a ratio of 1 – that is, perfect performance.

was observed, indicating an unequal distribution of response frequencies and therefore speaking against a random choice pattern. Most importantly, a significant interaction between rank and range [$F(5, 155) = 10.36$, $p < 0.001$] was observed, indicating that children indeed chose values close to the correct outcome. No other main effects or interaction were significant. The absence of significant interactions between rank and operation or between all three factors indicates that this response pattern was comparable for both arithmetic operations.

Did children's performance in the non-symbolic calculation task conform to Weber's law? We next examined how children responded to our different arithmetic problems. The left column of **Figure 4** shows the children's mean responses (chosen values) as a function of the size of the correct result, separately for the two operations. If the children were able to solve the arithmetic problems, the chosen value should increase as a function of the correct outcome. With increasing numerical magnitude, theory predicts an increasing variability of the chosen values (see the appendix in Barth et al., 2006). Finally, according to Weber's law, the increase in the chosen values should be paralleled by a proportional increase in response variability, as expressed in terms of their respective standard deviation, resulting in a constant coefficient of variation (CV, the ratio of the standard deviation, and mean of the subjects' responses) across arithmetic problems of different numerical magnitude. As can be seen in **Figure 4**, both children's mean responses (depicted as squares) and their standard deviation (depicted as circles) increased as a function of the correct outcome for both addition (black) and subtraction (gray). This impression was confirmed by repeated measures ANOVAs of mean and standard deviation with operation (addition, subtraction) and numerosity (8, 10, 19, 25) as within-group factors and age as between group factor. Mean responses [$F(3, 90) = 313.04$, $p < 0.001$, epsilon = 0.60] and variability [$F(3, 90) = 53.70$, $p < 0.001$, epsilon = 0.68] of responses increased significantly with increasing correct outcome. Interestingly, the increase of mean chosen value and its variation
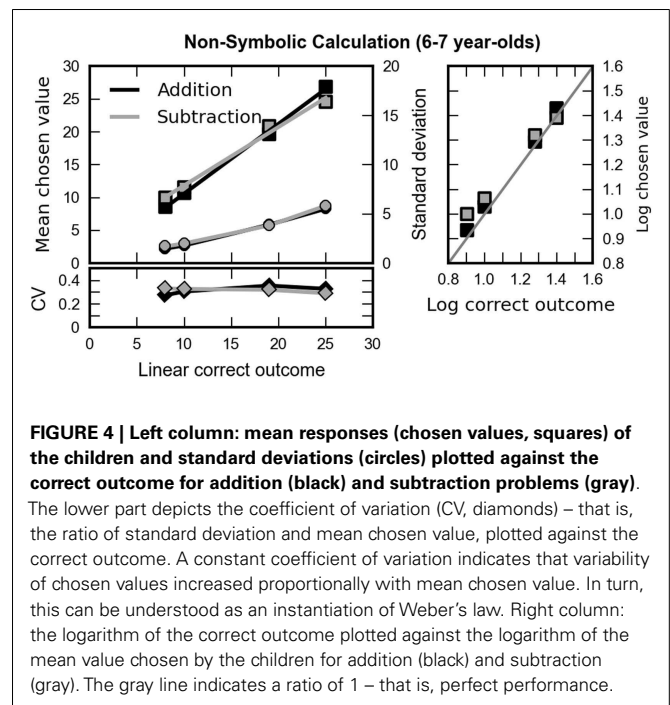
were stronger for addition than for subtraction as indicated by the significant interactions [mean: $F(3, 90) = 6.42$, $p = 0.001$, epsilon = 0.82; SD: $F(3, 90) = 3.12$, $p = 0.035$, epsilon = 0.91]. No other main effect or interaction was significant.

As can be seen in the lower left part of **Figure 4**, the CV was constant across the whole range of outcomes for addition and subtraction. This was tested statistically with a repeated measures ANOVA with operation (addition, subtraction) and numerosity (8, 10, 19, 25) as within-group factors and age as between group factor. No main effect or interaction reached statistical significance (minimum $p = 0.133$). To further corroborate this finding we calculated the difference between the correct outcome and the mean chosen value, once both of them had been transformed to a logarithmic scale, and calculated a repeated measures ANOVA on the standard deviations of these differences, with size of the correct result as the only factors, separately for both operations (addition and subtraction). Neither for addition [$F(3, 93) = 2.65$, $p = 0.053$] nor for subtraction [$F(3, 93) = 0.977$, $p = 0.407$] we observed a significant impact of problem size on the standard deviation of this index.

Taken together, these results suggest that data are well described by Weber's law which is in line with the assumption that the underlying mental magnitude representation is logarithmically compressed. Therefore, all following analyses concerning the OM effect were carried out in a logarithmic scale, using as input the difference between the logarithm of the correct outcome and the logarithm of the chosen value. Such analyses also have the advantage of more likely meeting the prerequisites of the ANOVA, which stipulates that all data have a fixed variability.

Did children show operational momentum in non-symbolic calculation? To quantify this OM effect, we computed a simple

estimate of response bias: the mean difference between the log of the subject's responses and the log correct result. This value was submitted to an ANOVA with operation as within-group factor and age as between group factor. Most importantly, no main effect of operation was observed [$F(1, 30) = 3.02$, $p = 0.092$], that is, no significant bias toward smaller responses for subtraction than for addition was observed for children. No significant interaction with age was observed [$F(1, 30) = 0.76$, $p = 0.39$]. Results are shown in **Figure 5**.

Cueing task
Overall performance in the cueing task was very good. Children committed only a total of 20 errors corresponding to 2% that were excluded from all subsequent analyses.
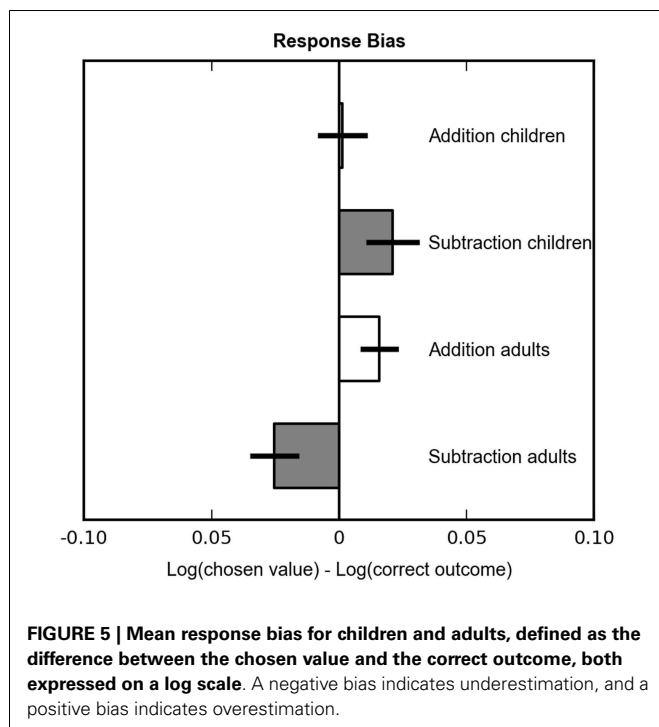
Mean reaction times from valid, invalid, and neutral conditions were computed per child and z-standardized per child (mean = 0, SD = 1) to account for high between-subject variability in reaction times from children.

Benefit and cost in the cueing task. First, we analyzed the effects of cueing on reaction times by computing the benefits and the costs of valid and invalid trials with respect the neutral condition, respectively. Adopting standard nomenclature from attention domain the benefit (neutral – valid) will be referred to as orienting effect and cost (neutral – invalid) will be referred to as reorienting effect. **Figure 6** (left) depicts the cueing effects and implies that the type of cue (valid, invalid, or neutral) had a measurable impact on children's performance. This impression was supported by a significant main effect of cue type in a repeated measures ANOVA with cue type (valid, invalid, neutral) and SOA (200 ms, 800 ms) as factors [$F(2, 30) = 21.13$, $p < 0.001$]. No other main effect or interaction was significant, implying that the observed cueing effects

were not statistically modulated by SOA. Paired t-tests revealed a significant orienting effect {valid trials were responded to faster than neutral trials [$t(15) = -2.72$, $p = 0.016$]} and a significant reorienting effect {neutral trials were responded to faster than invalid trials [$t(15) = -4.16$, $p = 0.001$]}.

Correlating cueing effects with calculation data
It has been argued that the OM effect is at least partially due to attentional shifts induced by the arithmetic operation that is operating on a spatially oriented mental magnitude representation. According to this account additions are associated with attentional shifts to the right and subtractions are associated with attentional shifts to the left. To test this account we assigned half of the children in the current study to a cueing paradigm. If the OM effect is a consequence of the interaction between the attentional systems and the mental magnitude system we should observe a correlation between both measures over children. That is, children with a large OM effect should also exhibit larger attentional cueing effects. To test this account we computed Pearson correlation coefficients between the OM bias (that is [$\log(\text{chosen}_{\text{Addition}}) - \log(\text{correct}_{\text{Addition}})$] − [$\log(\text{chosen}_{\text{Subtraction}}) - \log(\text{correct}_{\text{Subtraction}})$]) and the orienting and reorienting effects observed in the cueing paradigm. Note that this analysis, too, is based upon correct trials in the cueing paradigm only. While orienting did not correlate significantly with
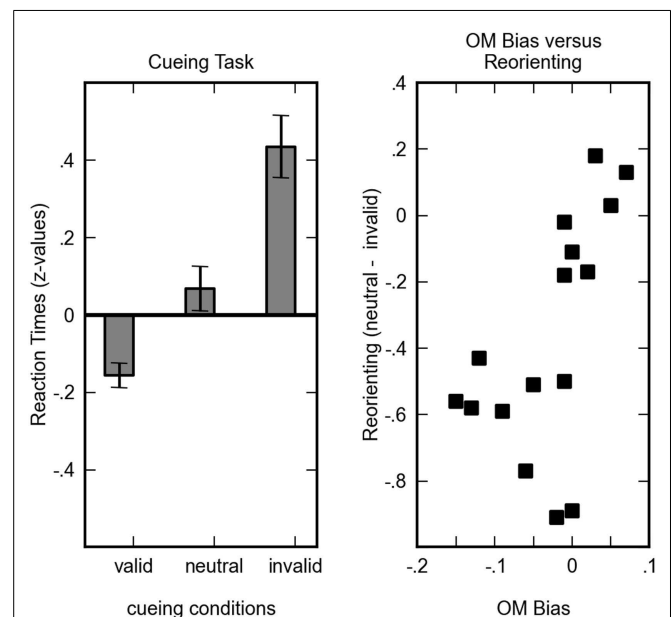
FIGURE 6 | Left column: *z*-standardized mean reaction times for valid, neutral and invalid conditions of the attention paradigm. Error bars indicate standard error of the mean. Right column: the reorienting effect (difference between neutral and invalid trials) plotted against the operational momentum bias. For reorienting better performance is indicated by numerically larger (i.e., less negative) values. A regular operational momentum effect corresponds to positive values, an inverse operational momentum effect corresponds to negative values. The correlation between reorienting and operational momentum signifies that the less children suffer from invalid cueing the more they are prone to exhibit a regular operational momentum effect.

FIGURE 5 | Mean response bias for children and adults, defined as the difference between the chosen value and the correct outcome, both expressed on a log scale. A negative bias indicates underestimation, and a positive bias indicates overestimation.

OM bias ($r = 0.14$, $p = 0.616$), a significant correlation between OM bias and reorienting was observed ($r = 0.59$, $p = 0.017$). The difference between invalid and valid trials did not significantly correlate with OM bias ($r = -0.37$, $p = 0.158$). In **Figure 6** (right) we plot the individual OM biases against the reorienting effect. It becomes evident that the relatively high correlation was not driven by few outliers but that over the entire range of reorienting and OM bias a higher OM bias was associated with smaller reorienting effects, that is with lower costs due to invalid cueing.

Correlating reading fluency with OM

We hypothesized that reading may corroborate existing spatial–numerical links and may thus be linked with OM. To test this idea we assessed individual reading capacities by measuring reading durations for a short text. Although reading durations showed substantial variability (mean = 114 s; SD = 94 s) no correlation was observed with OM [r(reading, OM bias) $= -0.18$, $p = 0.513$].

ADULTS

To test whether the current version of the paradigm is suited to reveal OM effects we administered the same paradigm to a group of 32 students from Barnard College. The same analysis steps as for children were performed for adults and described briefly below.

### Can participants memorize the operands?

Again we start by testing whether participants were able to correctly memorize the shown values. Both mean chosen values (6.7, 20.8, and 25.4 for displayed numerosities 6, 19, and 25, respectively) and the standard deviation of chosen memorized value increased significantly with presented numerosity [mean: $F(2, 62) = 473.14$, $p < 0.001$, epsilon = 0.83; mean: $F(2, 62) = 38.28$, $p < 0.001$] indicating that participants were capable of remembering the presented numerosities. In line with the assumption of the mental magnitude representation following Weber's law we observed a constant coefficient of variation that did not significantly covary with numerosity [$F(2, 60) = 1.33$, $p = 0.273$, epsilon = 0.88].

Did participants engage in approximate calculation or respond at random?

Next, we analyzed whether the subjects chose among the proposed choices at random. To this end we analyzed (arcsine-transformed) percentage of choice as the dependent variable in an ANOVA and rank of the subject's choice (one to six), range (second or fifth value correct), and operation (addition, subtraction) as factors. A main effect of rank [$F(5, 155) = 3.57$, $p = 0.006$, epsilon = 0.89] was observed, indicating an unequal distribution of response frequencies and therefore speaking against a random choice pattern. Most importantly, a significant interaction between rank and range [$F(5, 155) = 54.14$, $p < 0.001$] was observed, indicating that participants did not engage in a random choice pattern. Operation interacted significantly with range [$F(5, 155) = 4.22$, $p = 0.049$], rank [$F(5, 155) = 6.84$, $p < 0.001$, epsilon = 78], and with range and rank [$F(5, 155) = 9.88$, $p < 0.001$). Results are shown in **Figure 7**.

Did participant's performance in the non-symbolic calculation task conform to Weber's law?

We next examined how the subjects responded to the different arithmetic problems. Specifically, repeated measures ANOVAs
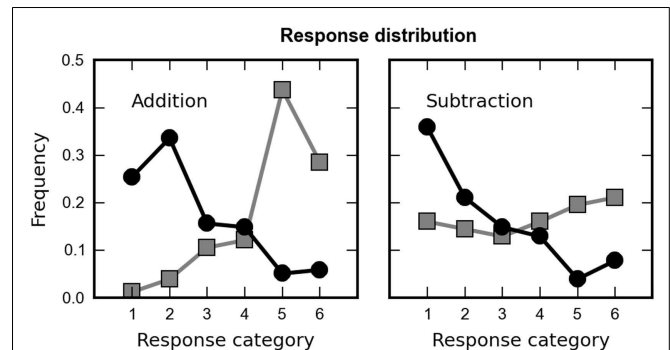
FIGURE 7 | Distribution of the participants' choices across the six proposed results, averaged over all arithmetic problems, separately for addition (left column) and subtraction (right column). Responses were not distributed randomly but depended on the range of response alternatives presented (high or low range shown as black squares and gray squares, respectively). Responses were centered around the values that were closest to the correct outcome (fifth for low range and second for high range).
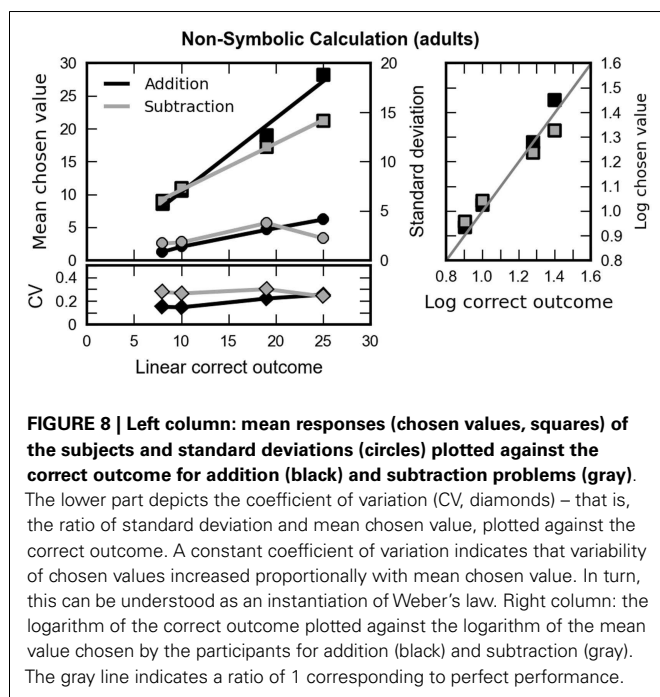
with numerosity and arithmetic operation revealed that: (a) chosen values increased as a function of correct outcome [$F(3, 93) = 764.36$, $p < 0.001$, epsilon = 0.85], (b) the variability of the choices increased with increasing correct outcome [$F(3, 93) = 73.19$, $p < 0.001$, epsilon = 0.86]. We found that the coefficient of variation increased with increasing correct outcome [$F(3, 93) = 3.82$, $p = 0.017$, epsilon = 0.87]. A significant interaction between operation and numerosity [$F(3, 93) = 6.15$, $p = 0.001$] was due to the fact that the increase was present only for addition [$F(3, 93) = 9.44$, $p < 0.001$] but not for subtraction [$F(3, 93) = 1.63$, $p = 0.188$]. This pattern of results was corroborated by the results of two repeated measures ANOVAs on the standard deviation of the difference between the chosen values and the correct outcomes after they had been transformed to log scale with size of the correct outcome as the only factor. We observed a significant main effect of size for addition problems [$F(3, 93) = 10.56$, $p < 0.001$] but the size of the outcome did not systematically influence response variability for subtraction [$F(3, 93) = 1.74$, $p = 0.164$]. Results are shown in **Figure 8**.

Did participants show operational momentum in non-symbolic calculation?

Finally, we analyzed the OM effect by computing the same bias as for children {[log(chosen value) − log(correct outcome)]} for each problem and averaged over addition and subtraction problems, separately. A paired sample t-test revealed a significant OM effect [$t(31) = 2.91$, $p = 0.007$] that took the form of a full cross-over effect. That is, participants significantly overestimated results for addition problems [$t(31) = 2.12$, $p = 0.042$] and under-estimated results for subtraction problems [$t(31) = -2.61$, $p = 0.014$]. Results are depicted in **Figure 5**.

JOINT ANALYSIS OF OM IN ADULTS AND CHILDREN

To statistically test the observed discrepancy of regular OM in adults and the absence of a statistically significant OM in children we submitted the log-scaled bias [log10(chosen value) − log10(correct outcome)] from both groups to a common

FIGURE 8 | Left column: mean responses (chosen values, squares) of the subjects and standard deviations (circles) plotted against the correct outcome for addition (black) and subtraction problems (gray). The lower part depicts the coefficient of variation (CV, diamonds) – that is, the ratio of standard deviation and mean chosen value, plotted against the correct outcome. A constant coefficient of variation indicates that variability of chosen values increased proportionally with mean chosen value. In turn, this can be understood as an instantiation of Weber's law. Right column: the logarithm of the correct outcome plotted against the logarithm of the mean value chosen by the participants for addition (black) and subtraction (gray). The gray line indicates a ratio of 1 corresponding to perfect performance.

ANOVA with type of arithmetic operation (addition vs. subtraction) as within-subjects factor and group (adults vs. children) as between-subjects factor. Type of arithmetic operation did not have a significant impact on the observed calculation bias [$F(1, 62) = 1.44$, $p = 0.234$]. No main effect of group was observed [$F(1, 62) = 2.601$, $p = 0.112$]. Most importantly, in line with the observed discrepancy type of arithmetic operation significantly interacted with group [$F(1, 62) = 11.62$, $p = 0.001$], statistically corroborating the observation that the differential impact of the arithmetic operation on the chosen values depends on the group. Adults show a differential impact of arithmetic operation while children tend not to. This is in line with the observation that the OM bias for addition and subtraction is negatively correlated over adults ($r = -0.343$, $p = 0.054$) but positively correlated over children ($r = 0.42$, $p = 0.017$). Put differently, those adults who tend to larger overestimation in addition also tend to larger underestimation in subtraction. In children this pattern is reversed. Children who tend to larger overestimation in addition also show larger overestimation in subtraction.

JOINT ANALYSIS OF OM IN ADULTS AND CHILDREN USING A BAYESIAN APPROACH

The repeated measures ANOVA model as implemented in SPSS assumes homoscedasticity, meaning that the variation of the dependent variable is the same for each experimental group and repeated measurement. In developmental research, this might however not be justified. If the comparison involves, for example, children, heteroscedasticity may be observed due to increased variation in functioning, compliance with the task, or both. Indeed, analyzing the standard deviation of the chosen values in a 2 (operation) × 4 (outcome) repeated measures ANOVA with age group (adults vs. children) as between group factor revealed a significant main effect of group [$F(1, 62) = 23.15$, $p < 0.001$], indicating that

adults responses were less variable than children's responses. Thus, the stipulated homoscedasticity cannot be assumed. As known from the statistical literature, heteroscedasticity can substantially decrease power (Wilcox et al., 1986; Wilcox, 1987), the probability of identifying an existing effect.

A straightforward way to overcome the implausible restriction of homoscedasticity is by using a Bayesian model with all (co)variances considered as unknown parameters. In the Bayesian approach, estimation is informed not only by information from the data (likelihood function), but also by *a priori* information (prior density function). Their product is proportional to the target of Bayesian estimation (posterior density). The posterior density is often computed with the help of Monte Carlo methods. For an introduction to the Bayesian approach for experimental researchers, see Kruschke and Safari Tech Books Online (2011).

Let $i = 1, 2, \ldots, 32$, $j = 1, 2$, $k = 1, 2$ index subjects (1 through 32), measurements (addition vs. subtraction), and groups (adults vs. children), respectively. Then the data form an $i \times j \times k$ array. We consider a multivariate normal likelihood. In probability notation, this can be put as:

$$y_{ik} \sim N(\mu_k, \Sigma_k), \qquad (1)$$

with the vector $y_{ik} = \begin{pmatrix} y_{i1k} \\ y_{i2k} \end{pmatrix}$ of the addition and subtraction scores of the ith subject from the kth group, $\mu_k = \begin{pmatrix} \mu_{1k} \\ \mu_{2k} \end{pmatrix}$ the kth group's means in addition and substraction, and $\Sigma_k = \begin{pmatrix} \sigma_{11k}^2 & \sigma_{12k}^2 \\ \sigma_{21k}^2 & \sigma_{22k}^2 \end{pmatrix}$ the group-specific covariance matrix. Notice that: (a) the model assumes correlated measurements as the secondary diagonal of Σ_k is not restricted to zero and (b) it assumes heteroscedasticity as the variances are allowed to be unequal across groups and measurements, why we may consider it as a heteroscedastic repeated measures ANOVA model. Prior densities must be specified to complete the model. We choose a flat normal density for the measurement means of each group because we wish to let the data dominate the analysis:

$$\mu_{jk} \sim N(0, 100000). \qquad (2)$$

For the same reason, we assign the following inverse Wishart density to the covariance matrices with identity matrix $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and 2 degrees of freedom:

$$\Sigma_k \sim inverse - Wishart(I, 2). \qquad (3)$$

Before running the model, data were standardized. We sampled μ_{1k}, μ_{2k}, σ_{11k}^2, σ_{12k}^2, and σ_{22k}^2 iteratively using JAGS 3.3.0. Convergence was checked by visual inspection. We discarded the first 100,000 out of 600,000 iterations as burn-in. Inference was based on the remaining. Marginal posterior densities including those of the effects of operation in adults and children, $\mu_{11} - \mu_{21}$ and $\mu_{12} - \mu_{22}$ respectively, and that of the interaction $(\mu_{11} - \mu_{21}) - (\mu_{12} - \mu_{22})$, are summarized in **Table 2**.

The probability that the effect is positive/negative in the light of the data is given by the appropriate integral of the

Table 2 | Marginal posterior densities from the Bayesian analysis.

	Mean	HDI[a]
MODEL PARAMETERS		
Means		
μ_{11}	0.226	[−0.054, 0.502]
μ_{21}	−0.513	[−0.863, −0.156]
μ_{12}	−0.034	[−0.391, 0.324]
μ_{22}	0.321	[−0.062, 0.702]
(Co)variances		
σ^2_{111}	0.635	[0.353, 0.974]
σ^2_{121}	−0.265	[−0.589, 0.031]
σ^2_{221}	1.026	[0.571, 1.576]
σ^2_{112}	1.056	[0.582, 1.617]
σ^2_{122}	0.460	[0.054, 0.930]
σ^2_{222}	1.204	[0.666, 1.846]
DERIVED QUANTITIES		
Effects		
$\mu_{11} - \mu_{21}$ (addition vs. subtraction adults)	0.739	[0.221, 1.254]
$\mu_{12} - \mu_{22}$ (addition vs. subtraction children)	−0.354	[−0.758, 0.048]
$(\mu_{11} - \mu_{21}) - (\mu_{12} - \mu_{22})$	1.093	[0.450, 1.756]

[a]The highest density interval is a credible region, including the most likely values.

marginal posterior density (Jackman, 2009). For $\mu_{11} - \mu_{21}$, $(\mu_{11} - \mu_{21}) - (\mu_{12} - \mu_{22})$, and $\mu_{12} - \mu_{22}$ this probability is 0.997, 0.999, and 0.958, respectively, 0.003, 0.001, and 0.042 are the probabilities of the contrary, namely that the effect is zero or less/more than zero. Considering their ratios, we conclude that beside decisive support for OM in adults and a difference in OM between adults and children, our data strongly supports the idea of an inverse OM in children.

DISCUSSION
In the current study we administered a non-symbolic calculation paradigm – along with a reading and an attentional cueing task – to 6- and 7-year olds to examine the presence and determinants of OM. Four main findings are of note. First, we replicate the presence of proficient non-symbolic addition and subtraction in an adult population, along with a systematic overestimation of addition outcomes, and underestimation of subtraction outcomes (McCrink et al., 2007; Knops et al., 2009b). Second, the children tested here were capable of non-symbolic addition and subtraction, using only their "number sense"; they reliably altered their responses to the offered outcomes to correspond with a mental calculation of the estimated correct outcomes, and did so without the aid of confounding perceptual cues. Third, while a group of adult controls showed a regular OM effect, children did not show a significant OM effect using classical statistical NHST. However, using Bayesian inference we observed a significant inverse OM effect. That is, subtraction problems lead to significantly larger overestimations than addition problems. Finally, there was a relationship between a child's level of attentional reorienting and their propensity to exhibit regular OM; the lower the cost of the invalid cue the more regular OM bias exhibited by that child. Counter to hypotheses that offer self-directed automaticity of reading as a driver of spatial–numerical links, we found no relationship between reading fluency and OM.

What do the current findings implicate for the different hypotheses concerning the basis of OM and the developmental trajectory of OM?

THE COMPRESSION HYPOTHESIS
The compression hypothesis assumes that the OM results from a flawed uncompression operation during the course of manipulating mental magnitudes (McCrink et al., 2007; Chen and Verguts, 2012). According to this hypothesis a regular OM effect was expected in the tested age range. The observed absence of an overall OM bias under classical statistical NHST in combination with a significantly reversed OM under Bayesian approach is hard to reconcile with this notion. One might argue that the compression of the MNL is not identical for adults and children at the age of 6–7 years, and therefore the OM should differ between adults and children. However, if anything compression of the mental magnitude representation is *more* pronounced for children in the tested age range, implying a regular OM bias that is even more pronounced than in adulthood (Siegler and Opfer, 2003; Opfer and Siegler, 2007). For example, Berteletti et al. (2012) found no significant differences between linear and logarithmic model for number to position task in first (mean age: 6;11) and second grade (mean age: 7;11). Preschoolers (mean age: 5;8) were best fit by logarithmic and as of third grade (mean age: 8;9) linear models provided best fit. Hence, the shift from logarithmic to linear mapping of numbers to positions occurs only in second or third grade, when children are older than the sample tested here. Therefore, the present results speak against a flawed compression-uncompression mechanism as the driving factor of the OM bias.

THE HEURISTICS APPROACH
The heuristics approach (McCrink and Wynn, 2009, p. 407) suggests that children deploy a general arithmetic principle of "if adding, accept more" than the original operand and "if subtracting, accept less." According to this approach, too, an OM bias was expected for the tested age range (that is, given an addition and subtraction problem that yield the same objective answer, the average subjective outcome chosen as correct for addition will be higher than that chosen for subtraction). Moreover, the response distribution was expected to differ significantly from the distribution observed for adults. Under the strictest interpretation of this theory, if children had adopted such a heuristic rather than engaging in an approximate calculation they would have frequently chosen results that are larger than the first operand in addition and smaller than the first operand in subtraction, with no differentiation between somewhat vs. extremely larger/smaller. This would result in a response distribution that plateaus at results discriminably larger than the initial outcome and ranging up (addition) or those that start at any outcome discriminably smaller than the initial operand and go down (subtraction). The present study yielded two findings that speak against this heuristic approach. First, we did not observe an overall OM bias in 6- and 7-year-old children using NHST in combination with a significant inverse OM under Bayesian approach. Second, the response distributions did not follow the expected pattern under the assumption of a pure heuristic. Rather, the distributions largely resembled those observed for adult participants, with distinct response peaks (albeit less pronounced for subtraction problems). For example, for the high response

range in addition problems the observed modal value was actually numerically smaller than the actual outcome. Together, these results imply that children did indeed engage in approximate calculation and speaks against the assumed heuristic of accepting generically "more" with addition or "less" with subtraction.

READING FLUENCY ACCOUNT

The hypothesis that reading fluency underlies the formation of a spatial–numerical link, and its resultant OM bias, was also unsupported. Although there was a wide range of reading ability (ranging between 36 and 320 s to read through a short vignette) this factor did not correlate with a propensity to show regular OM. While numerous studies have found that the reading directionality of adults clearly modulates the traditional SNARC effect, it is likely not responsible for *instantiating* it (Shaki et al., 2009, 2012). In addition to the current findings, there are several studies which show the presence of other spatial–numerical relationships before the onset of reading (Opfer et al., 2010; Berteletti et al., 2012; Shaki et al., 2012). This suggests that reading is not the driving factor in the formation of spatial–numerical links. Other aspects of the cultural milieu may lead to the development of spatial–numerical links (such as seeing adults model directional counting (Opfer et al., 2010), or utilize gesture in a culturally consistent fashion) which subsequently may be modulated by reading fluency.

ATTENTIONAL SHIFTS ACCOUNT

The attentional shift account explains OM as the result of shifts of spatial attention along the MNL that lead participants to prefer outcomes in the "direction" of the arithmetic operation (Knops et al., 2009a). This account predicts that children exhibit a response distribution similar to the pattern observed in adults and a relationship between attentional indices and the strength and/or presence of OM. Children showed an inverse OM bias under a Bayesian approach which is not in line with the predictions of the attentional shift hypothesis. The OM bias was correlated with the reorienting effect. With decreasing reorientation effect the tendency to exhibit a regular OM bias increased across children. Reorientation captures the ability to switch attention from invalidly cued locations to the uncued location at which the actual stimulus appears (Carrasco, 2011). Children who are more proficient in this process tend to show a more adult-like OM bias.

On the neural level in adults reorienting has been associated with joint activation in two distinct but intertwined cortical systems (Corbetta et al., 2008), the ventral attention system (VAS) and the dorsal attention system (DAS). The VAS encompasses right inferior parietal regions around the temporo-parietal junction and ventral frontal cortex including parts of middle frontal gyrus, inferior frontal gyrus, Insula, and frontal operculum (Corbetta et al., 2008). The DAS comprises bilateral areas in the intraparietal sulcus, superior parietal cortex, and frontal eye fields. The DAS is associated with goal-directed orienting of attention that biases the processing of relevant stimuli. In contrast, the VAS is activated by salient but unexpected stimuli and has been proposed to be suppressed during periods of focused attention to prevent reorienting to distracting events. The VAS likely receives filtering information about whether or not an unexpected stimulus is salient from prefrontal cortex (Shulman et al., 2003). Activity in the VAS is modulated at the transition point between two tasks or two stimuli

(Dosenbach et al., 2006). Against this background, the size of the reorienting effect might be interpreted as an index for the integrity of the described attentional systems, and/or the extent to which the attentional system is connected to executive control functions (situated in prefrontal cortex). Both reorienting and executive control have been shown to be functional yet immature in 6- to 7-year olds (Konrad et al., 2005; Wetzel et al., 2006, 2009; Carp et al., 2012). Thus children who show a smaller reorientation effect might be more mature in developing situational control over distracting stimuli, and therefore more likely to exhibit an adult-like OM. Together with a functional attentional selection system (DAS) – as evidenced by the significant benefit from valid cues – this implies a key role for a functional and mature attentional system for the OM to arise in the context of non-symbolic calculation tasks. Surprisingly, OM propensity did not correlate with orienting ($r = 0.14$) as predicted by the attentional shift hypothesis. We can only speculate about the reasons for this finding. A regular OM effect may rely on automatic and reliable number-space associations and a full-fledged attentional system. Under this assumption we would speculate to observe increasing OM bias with increasing age. Second, we may hypothesize that correlation between OM and attention increases with age. For the moment these considerations are speculative and need to be addressed in further studies. In this respect the present study presents the first pieces of evidence for the developmental trajectory of the OM effect and its relation to other cognitive domains such as reading and attention. The observed pattern of results is not fully compatible with the attentional shift hypothesis. Nevertheless, the observed response pattern in combination with the significant correlation between reorienting and OM make this hypothesis a promising theoretical approach to delineate the developmental trajectory of the OM effect.

In sum, in the present study we tested several existing hypotheses about the origin of the OM effect by administering a non-symbolic calculation task to children aged between 6 and 7 years. Most crucially, using a Bayesian framework we observed a significant inverse OM effect in children. Testing a sample of college students with the same paradigm revealed a significant regular OM. The children's results are hard to reconcile with the proposed theoretical accounts for the OM, namely the attentional shift hypothesis, the compression hypothesis, and the heuristic account. The propensity to show a regular OM effect correlated with reorienting scores in an adapted Posner paradigm, linking the OM to the attentional system. We believe that the current child-friendly paradigm offers a promising avenue to further explore the development of spatial–numerical links, and that these findings lead to novel predictions based on the relationship between distinct attention systems, space, and number in both children and adults.

REFERENCES

Bächtold, D., Baumuller, M., and Brugger, P. (1998). Stimulus-response compatibility in representational space. *Neuropsychologia* 36, 731–735. doi:10.1016/S0028-3932(98)00002-5

Barth, H., La, M. K., Lipton, J., and Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14116–14121. doi:10.1073/pnas.0505512102

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., and Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition* 98, 199–222. doi:10.1016/j.cognition.2004.09.011

Barth, H. C., and Paladino, A. M. (2011). The development of numerical estimation: evidence against a representational shift. *Dev. Sci.* 14, 125–135. doi:10.1111/j.1467-7687.2010.00962.x

Berch, D. B., Foley, E. J., Hill, R. J., and Ryan, P. M. (1999). Extracting parity and magnitude from Arabic numerals: developmental changes in number processing and mental representation. *J. Exp. Child Psychol.* 74, 286–308. doi:10.1006/jecp.1999.2518

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., and Zorzi, M. (2010). Numerical estimation in preschoolers. *Dev. Psychol.* 46, 545–551. doi:10.1037/a0017887

Berteletti, I., Lucangeli, D., and Zorzi, M. (2012). Representation of numerical and non-numerical order in children. *Cognition* 124, 304–313. doi:10.1016/j.cognition.2012.05.015

Brannon, E. M., Wusthoff, C. J., Gallistel, C. R., and Gibbon, J. (2001). Numerical subtraction in the pigeon: evidence for a linear subjective number scale. *Psychol. Sci.* 12, 238–243. doi:10.1111/1467-9280.00342

Calabria, M., and Rossetti, Y. (2005). Interference between number processing and line bisection: a methodology. *Neuropsychologia* 43, 779–783. doi:10.1016/j.neuropsychologia.2004.06.027

Carp, J., Fitzgerald, K. D., Taylor, S. F., and Weissman, D. H. (2012). Removing the effect of response time on brain activity reveals developmental differences in conflict processing in the posterior medial prefrontal cortex. *Neuroimage* 59, 853–860. doi:10.1016/j.neuroimage.2011.07.064

Carrasco, M. (2011). Visual attention: the past 25 years. *Vision Res.* 51, 1484–1525. doi:10.1016/j.visres.2011.04.012

Chen, Q., and Verguts, T. (2012). Spatial intuition in elementary arithmetic: a neurocomputational account. *PLoS ONE* 7:e31180. doi:10.1371/journal.pone.0031180

Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324. doi:10.1016/j.neuron.2008.04.017

Cordes, S., Williams, C. L., and Meck, W. H. (2007). Common representations of abstract quantities. *Curr. Dir. Psychol. Sci.* 16, 156–161. doi:10.1111/j.1467-8721.2007.00495.x

Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* 122, 371–396. doi:10.1037/0096-3445.122.3.371

Dehaene, S., and Changeux, J. P. (1993). Development of elementary numerical abilities – a neuronal model. *J. Cogn. Neurosci.* 5, 390–407. doi:10.1162/jocn.1993.5.4.390

Dehaene, S., Izard, V., Spelke, E., and Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science* 320, 1217–1220. doi:10.1126/science.1156540

Dosenbach, N. U., Visscher, K. M., Palmer, E. D., Miezin, F. M., Wenger, K. K., Kang, H. C., et al. (2006). A core system for the implementation of task sets. *Neuron* 50, 799–812. doi:10.1016/j.neuron.2006.04.031

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., and Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year old children: evidence for a segmented linear model. *J. Exp. Child Psychol.* 99, 1–17. doi:10.1016/j.jecp.2007.08.006

Fischer, M. H. (2001). Number processing induces spatial performance biases. *Neurology* 57, 822–826. doi:10.1212/WNL.57.5.822

Fischer, M. H., Castel, A. D., Dodd, M. D., and Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nat. Neurosci.* 6, 555–556. doi:10.1038/nn1066

Freyd, J. J., and Finke, R. A. (1984). Representational momentum. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 126–132. doi:10.1037/0278-7393.10.1.126

Gallistel, C. R. (1990). *The Organization of Learning.* Cambridge, MA: MIT Press.

Gallistel, C. R. (2011). "Mental magnitudes," in *Space, Time and Number in the Brain*, eds S. Dehaene and E. M. Brannon (New York: Associated Press), 3–12.

Gallistel, C. R., and Gelman, I. I. (2000). Non-verbal numerical cognition: from reals to integers. *Trends Cogn. Sci. (Regul. Ed.)* 4, 59–65. doi:10.1016/S1364-6613(99)01424-2

Huynh, H., and Feldt, L. S. (1970). Conditions under which mean square ratios in repeqted measurements designs have exact F-distributions. *J. Am. Stat. Assoc.* 65, 1582–1589. doi:10.1080/01621459.1970.10481187

Imbo, I., Brauwer, J. D., Fias, W., and Gevers, W. (2012). The development of the SNARC effect: evidence for early verbal coding. *J. Exp. Child Psychol.* 111, 671–680. doi:10.1016/j.jecp.2011.09.002

Izard, V., and Dehaene, S. (2008). Calibrating the mental number line. *Cognition* 106, 1221–1247. doi:10.1016/j.cognition.2007.06.004

Jackman, S. (2009). *Bayesian Analysis for the Social Sciences.* Hoboken, NJ: Wiley.

Karolis, V., Iuculano, T., and Butterworth, B. (2011). Mapping numerical magnitudes along the right lines: differentiating between scale and bias. *J. Exp. Psychol. Gen.* 140, 693–706. doi:10.1037/a0024255

Knops, A., Thirion, B., Hubbard, E. M., Michel, V., and Dehaene, S. (2009a). Recruitment of an area involved in eye movements during mental arithmetic. *Science* 324, 1583–1585. doi:10.1126/science.1171599

Knops, A., Viarouge, A., and Dehaene, S. (2009b). Dynamic representations underlying symbolic and nonsymbolic calculation: evidence from the operational momentum effect. *Atten. Percept. Psychophys.* 71, 803–821. doi:10.3758/APP.71.4.803

Konrad, K., Neufang, S., Thiel, C. M., Specht, K., Hanisch, C., Fan, J., et al. (2005). Development of attentional networks: an fMRI study with children and adults. *Neuroimage* 28, 429–439. doi:10.1016/j.neuroimage.2005.06.065

Kruschke, J. K., and Safari Tech Books Online. (2011). *Doing Bayesian Data Analysis a Tutorial with R and BUGS.* Amsterdam: Academic Press.

McCrink, K., Dehaene, S., and Dehaene-Lambertz, G. (2007). Moving along the number line: operational momentum in nonsymbolic arithmetic. *Percept. Psychophys.* 69, 1324–1333. doi:10.3758/BF03192949

McCrink, K., and Wynn, K. (2004). Large-number addition and subtraction by 9-month-old infants. *Psychol. Sci.* 15, 776–781. doi:10.1111/j.0956-7976.2004.00755.x

McCrink, K., and Wynn, K. (2009). Operational momentum in large-number addition and subtraction by 9-month-olds. *J. Exp. Child Psychol.* 103, 400–408. doi:10.1016/j.jecp.2009.01.013

Moyer, R. S., and Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature* 215, 1519–1520. doi:10.1038/2151519a0

Nicholls, M. E., Loftus, A. M., and Gevers, W. (2008). Look, no hands: a perceptual task shows that number magnitude induces shifts of attention. *Psychon. Bull. Rev.* 15, 413–418. doi:10.3758/PBR.15.2.413

Nieder, A., and Dehaene, S. (2009). Representation of number in the brain. *Annu. Rev. Neurosci.* 32, 185–208. doi:10.1146/annurev.neuro.051508.135550

Nieder, A., and Miller, E. K. (2003). Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* 37, 149–157. doi:10.1016/S0896-6273(02)01144-3

Opfer, J. E., and Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cogn. Psychol.* 55, 169–195. doi:10.1016/j.cogpsych.2006.09.002

Opfer, J. E., Thompson, C. A., and Furlong, E. E. (2010). Early development of spatial-numeric associations: evidence from spatial and quantitative performance of preschoolers. *Dev. Sci.* 13, 761–771. doi:10.1111/j.1467-7687.2009.00934.x

Patro, K., and Haman, M. (2012). The spatial-numerical congruity effect in preschoolers. *J. Exp. Child Psychol.* 111, 534–542. doi:10.1016/j.jecp.2011.09.006

Pinhas, M., and Fischer, M. H. (2008). Mental movements without magnitude? A study of spatial biases in symbolic arithmetic. *Cognition* 109, 408–415. doi:10.1016/j.cognition.2008.09.003

R Development Core Team. (2010). *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing.

Shaki, S., Fischer, M. H., and Gobel, S. M. (2012). Direction counts: a comparative study of spatially directional counting biases in cultures with different reading directions. *J. Exp. Child Psychol.* 112, 275–281. doi:10.1016/j.jecp.2011.12.005

Shaki, S., Fischer, M. H., and Petrusic, W. M. (2009). Reading habits for both words and numbers

contribute to the SNARC effect. *Psychon. Bull. Rev.* 16, 328–331. doi:10.3758/PBR.16.2.328

Shulman, G. L., McAvoy, M. P., Cowan, M. C., Astafiev, S. V., Tansy, A. P., D'Avossa, G., et al. (2003). Quantitative analysis of attention and detection signals during visual search. *J. Neurophysiol.* 90, 3384–3397. doi:10.1152/jn.00343.2003

Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–243. doi:10.1111/1467-9280.02438

Stoianov, I., and Zorzi, M. (2012). Emergence of a "visual number sense" in hierarchical generative models. *Nat. Neurosci.* 15, 194–196. doi:10.1038/nn.2996

van Galen, M. S., and Reitsma, P. (2008). Developing access to number magnitude: a study of the SNARC effect in 7- to 9-year-olds. *J. Exp. Child Psychol.* 101, 99–113. doi:10.1016/j.jecp.2008.05.001

Viarouge, A., Hubbard, E. M., Dehaene, S., and Sackur, J. (2010). Number line compression and the illusory perception of random numbers. *Exp. Psychol.* 57, 1–9. doi:10.1027/1618-3169/a000055

Wetzel, N., Widmann, A., Berti, S., and Schroger, E. (2006). The development of involuntary and voluntary attention from childhood to adulthood: a combined behavioral and event-related potential study. *Clin. Neurophysiol.* 117, 2191–2203. doi:10.1016/j.clinph.2006.06.717

Wetzel, N., Widmann, A., and Schroger, E. (2009). The cognitive control of distraction by novelty in children aged 7-8 and adults. *Psychophysiology* 46, 607–616. doi:10.1111/j.1469-8986.2009.00789.x

Wilcox, R. R. (1987). A heteroscedastic ANOVA procedure with specified power. *J. Educ. Behav. Stat.* 12, 271–281. doi:10.2307/1164687

Wilcox, R. R., Charlin, V. L., and Thompson, K. L. (1986). New monte carlo results on the robustness of the ANOVA f, w and f statistics. *Commun. Stat. Simul. Comput.* 15, 933–943. doi:10.1080/03610918608812553

Wynn, K. (1992). Addition and subtraction by human infants. *Nature* 358, 749–750. doi:10.1038/358749a0

Zorzi, M., Priftis, K., and Umilta, C. (2002). Brain damage: neglect disrupts the mental number line. *Nature* 417, 138–139. doi:10.1038/417138a

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Operational momentum effect in children with and without developmental dyscalculia

Karin Kucian[1,2]*, Fabienne Plangger[1], Ruth O'Gorman[1,2,3] and Michael von Aster[1,2,4,5]

[1] Center for MR-Research, University Children's Hospital Zurich, Zurich, Switzerland
[2] Children's Research Center, University Children's Hospital Zurich, Zurich, Switzerland
[3] Center of Integrative Human Physiology, University of Zurich, Switzerland
[4] Department of Child- and Adolescent Psychiatry, DRK-Hospital Westend Berlin, Germany
[5] Clinical Psychology and Psychotherapy, University of Potsdam, Potsdam, Germany
*Correspondence: karin.kucian@kispi.uzh.ch

A commentary on

Examining the presence and determinants of operational momentum in childhood
by Knops, A., Zitzmann, S., and McCrink, K. (2013). Front. Psychol. 4:325. doi: 10.3389/fpsyg.2013.00325

THE PRESENCE OF OPERATIONAL MOMENTUM IN CHILDHOOD

In their valuable article, Knops et al. (2013) challenge the existence of the operational momentum effect (OME) in children. The OME is characterized by the tendency to overestimate the result of addition problems and to underestimate the result of subtractions (McCrink et al., 2007). In line with previous findings, they replicated the OME in adults using a non-symbolic approximation task. In contrast, children did not exhibit such an OME. This finding was quite unexpected since current studies claim that the OME is present in childhood, as early as 9 months of age (Pinhas and Fischer, 2008; McCrink and Wynn, 2009). Together with the evaluation of attentional orienting capacity, the authors concluded that an attentional shift along the mental number line (MNL) most probably explains the OME.

We take the liberty of adding own results in this commentary which further support these findings. We have also tested the OME in typically achieving children and a matched group with developmental dyscalculia (DD) (**Table 1**). Typically achieving children did not show an OME in a symbolic number line task, as children underestimated the location of results for both additions and subtractions on the number line (**Figure 1**). In contrast to the study demonstrating an OME in 9-month old babies (McCrink and Wynn, 2009), we used a symbolic numerical task similar to the one used by Pinhas and Fischer (2008) who observed a reliable OME. However, their participants were already in adolescence. Therefore, it might be possible that since school-age children have lower experience in symbolic processing of calculations, an unconscious shift of attention on the MNL becomes evident only with increasing expertise and automatization.

Together with the findings of Knops et al., it seems that the left–right associations underlying the OME are dependent on development and experience. One might hypothesize that a complex interaction between visuo-spatial and attentional processes together with number related skills influence its development whereupon an early predisposition to relate representations of non-symbolic numerical magnitude to spatial length builds a core system (De Hevia and Spelke, 2010). In combination with cultural conventions such as reading direction, the experience of specific left-small/right-large associations (Berch et al., 1999; Opfer et al., 2010), numeric magnitude and number line estimation (Siegler and Booth, 2004) and the understanding of arithmetic concepts might lead first to the OME by non-symbolic presentation and later, after the acquisition of the symbolic number system, to the OME found by symbolic presentation (Pinhas and Fischer, 2008). That the OME is possibly subject to the dynamic nature of developing math proficiency would also be in line with current neuropsychological (Von Aster and Shalev, 2007) and neuronal models (Kucian and Kaufmann, 2009) suggesting the development of MNL representation successively depending on previous processes of representing numerical magnitudes by first verbal and later Arabic symbols. However, future studies will have to address these hypothesis systematically.

According to the assumption of such a hierarchical and interwoven model of OME, the impairment of one or several aspects may affect the development of the OME negatively.

THE OME IN CHILDREN WITH DEVELOPMENTAL DYSCALCULIA

DD is a specific learning disability of number processing and calculation. In addition to profound problems in numerical understanding, abnormalities in visuo-spatial, and attentional processes have also been associated with DD.

As expected, no OME was observed in our tested group of children with DD (**Figure 1**). Therefore, one might speculate that a lack of numerical understanding and reduced visuo-spatial and attentional functions, as often found in children with DD, might hinder the development of an OME.

LIMITATIONS

It is important to note that an adult control group is missing in our study. However,

Table 1 | Demographic and testmetric data.

	Dyscalculic children	Control children	*p*-value
N (f/m)	16 (10/6)	16 (9/7)	
Handedness (right/ambidexter/left)	14/1/1	12/0/4	
Age (years)	9.5 (0.8)	9.5 (1.1)	0.986
School grade	3.1 (1.1)	2.9 (0.9)	0.604
General IQ	99.0 (6.9)	110.1 (7.0)	0.000
Performance IQ	100.0 (12.2)	108.6 (10.2)	0.038
Verbal IQ	98.3 (6.8)	111.1 (6.7)	0.000
Numeracy (percentile rank)	14.7 (20.1)	68.9 (18.8)	0.000
Visuo-spatial memory span (block span)	4.6 (0.7)	4.9 (1.0)	0.361
Visuo-spatial working memory (block span)	2.1 (0.7)	2.8 (0.9)	0.043

General IQ was assessed by the mean of following subtests of the WISC-III (Wechsler, 1999): similarities, picture arrangement, calculation, block design, vocabulary. Performance IQ, mean of picture arrangement, calculation, block design; Verbal IQ, mean of similarities, vocabulary. Numeracy was measured by the neuropsychological test battery for number processing and calculation in children ZAREKI-R (Von Aster et al., 2006). Visuo-spatial memory span was determined by the Corsi-Block-Tapping test (Corsi, 1972) and visuo-spatial working memory by the Corsi-Block-Suppression test (Beblo et al., 2004).
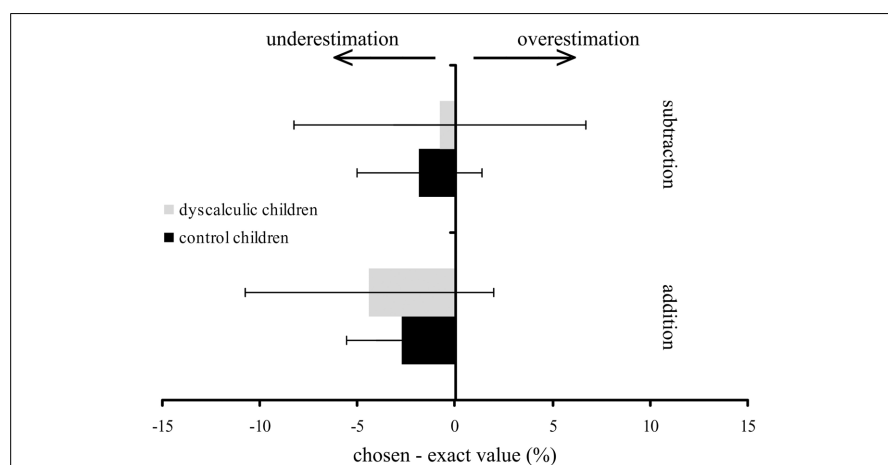
FIGURE 1 | Operational momentum effect in children with developmental dyscalculia and controls. Mean difference between chosen and exact value on a number line from 0–100 of children with (gray) and without DD (black). Negative values indicate an underestimation and positive values an overestimation of the position on the number line.

since Pinhas and Fischer (2008) demonstrated an OME in adolescents by means of a comparable paradigm to ours, we assume that the lack of OME in children is not due to differences in task design.

DOES A FLAWED UNCOMPRESSION OF NUMERICAL INFORMATION CAUSE THE OME?

Knops et al. argue that the lack of an OME during childhood speaks against a flawed compression-uncompression mechanism as the driving factor of the operational momentum bias, which implies that the OME is caused by a systematic bias during uncompression of a logarithmic representation of the MNL. Furthermore, number

representations change from a logarithmic mapping to a linear mapping scheme during development and familiarity in a number range (Siegler and Opfer, 2003). Hence, the MNL representation cannot be assumed to be identical in children and adults examined by Knops. However, as outlined by the authors, if a flawed uncompression of numerical information causes the OME, it should be more pronounced in children who still represent numbers in a logarithmic fashion.

In contrast to the cohort of Knops et al., children in our study were older and showed a linear function describing the MNL representation like in adults, but exhibited no OME. Moreover, our

results indicated that even after the completion of a specific number line training (Kucian et al., 2011), no OME was evident, although the training had a positive effect on the MNL representation. Therefore, our results further corroborate that differences in uncompression mechanisms of the MNL are unlikely to solely cause the OME.

CONCLUSIONS

Findings reported by Knops et al. and our study corroborate that the OME is not necessarily present in childhood and unlikely to be caused by flawed uncompression of the MNL. In addition, our results further point to a possible negative impact of DD on the development of the OME. In conclusion, the OME is probably dependent on development and a complex interaction of the maturity of numerical skills, visuo-spatial, and attentional processes as well as cultural conventions.

REFERENCES

Beblo, T., Macek, C., Brinkers, I., Hartje, W., and Klaver, P. (2004). A new approach in clinical neuropsychology to the assessment of spatial working memory: the block suppression test. *J. Clin. Exp. Neuropsychol.* 26, 105–114. doi: 10.1076/jcen.26.1.105.23938

Berch, D. B., Foley, E. J., Hill, R. J., and Mcdonough Ryan, P. (1999). Extracting parity and magnitude from arabic numerals: developmental changes in number processing and mental representation. *J. Exp. Child Psychol.* 74, 286–308. doi: 10.1006/jecp.1999.2518

Corsi, P. M. (1972). Human memory and the temporal region of the brain. *Diss. Abstr. Int.* 34, 891.

De Hevia, M. D., and Spelke, E. S. (2010). Number-space mapping in human infants. *Psychol. Sci.* 21, 653–660. doi: 10.1177/0956797610366091

Knops, A., Zitzmann, S., and Mccrink, K. (2013). Examining the presence and determinants of operational momentum in childhood. *Front. Psychol.* 4:325. doi: 10.3389/fpsyg.2013.00325

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schonmann, C., Plangger, F., et al. (2011). Mental number line training in children with developmental dyscalculia. *Neuroimage* 57, 782–795. doi: 10.1016/j.neuroimage.2011.01.070

Kucian, K., and Kaufmann, L. (2009). A developmental model of number representation. *J. Behav. Brain Sci.* 32, 340–341. doi: 10.1017/S0140525X09990069

McCrink, K., Dehaene, S., and Dehaene-Lambertz, G. (2007). Moving along the number line: operational momentum in nonsymbolic arithmetic. *Percept. Psychophys.* 69, 1324–1333. doi: 10.3758/BF03192949

McCrink, K., and Wynn, K. (2009). Operational momentum in large-number addition and subtraction by 9-month-olds. *J. Exp. Child Psychol.* 103, 400–408. doi: 10.1016/j.jecp.2009.01.013

Opfer, J. E., Thompson, C. A., and Furlong, E. E. (2010). Early development of spatial-numeric associations: evidence from spatial and quantitative performance of preschoolers. *Dev. Sci.* 13, 761–771. doi: 10.1111/j.1467-7687.2009.00934.x

Pinhas, M., and Fischer, M. H. (2008). Mental movements without magnitude? A study of spatial biases in symbolic arithmetic. *Cognition* 109, 408–415. doi: 10.1016/j.cognition.2008.09.003

Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi: 10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–243. doi: 10.1111/1467-9280.02438

Von Aster, M., and Shalev, R. (2007). Number development and developmental dyscalculia. *Dev. Med. Child Neurol.* 49, 868–873. doi: 10.1111/j.1469-8749.2007.00868.x

Von Aster, M., Weinhold Zulauf, M., and Horn, R. (2006). *ZAREKI-R (Neuropsychological Test Battery for Number Processing and Calculation in Children), revidierte Version*. Frankfurt: Harcourt Test Services.

Wechsler, D. (1999). *WISC-III Wechsler Intelligence Scale for Children*. Bern, Göttingen, Toronto, Seattle: Hans Huber.

Unbounding the mental number line—new evidence on children's spatial representation of numbers

Tanja Link[1]*, Stefan Huber[2], Hans-Christoph Nuerk[1,2] and Korbinian Moeller[1,2]

[1] Department of Psychology, Eberhard Karls University, Tuebingen, Germany
[2] Knowledge Media Research Center, Tuebingen, Germany

Number line estimation (i.e., indicating the position of a given number on a physical line) is a standard assessment of children's spatial representation of number magnitude. Importantly, there is an ongoing debate on the question in how far the bounded task version with start and endpoint given (e.g., 0 and 100) might induce specific estimation strategies and thus may not allow for unbiased inferences on the underlying representation. Recently, a new unbounded version of the task was suggested with only the start point and a unit fixed (e.g., the distance from 0 to 1). In adults this task provided a less biased index of the spatial representation of number magnitude. Yet, so far there are no children data available for the unbounded number line estimation task. Therefore, we conducted a cross-sectional study on primary school children performing both, the bounded and the unbounded version of the task. We observed clear evidence for systematic strategic influences (i.e., the consideration of reference points) in the bounded number line estimation task for children older than grade two whereas there were no such indications for the unbounded version for any one of the age groups. In summary, the current data corroborate the unbounded number line estimation task to be a valuable tool for assessing children's spatial representation of number magnitude in a systematic and unbiased manner. Yet, similar results for the bounded and the unbounded version of the task for first- and second-graders may indicate that both versions of the task might assess the same underlying representation for relatively younger children—at least in number ranges familiar to the children assessed. This is of particular importance for inferences about the nature and development of children's magnitude representation.

Keywords: mental number line, number line estimation, estimation strategies, proportion judgment, numerical development

INTRODUCTION

The metaphor of a mental number line (Moyer and Landauer, 1967; Restle, 1970) describing the (spatial) representation of number magnitude is widely recognized (for overviews see Hubbard et al., 2005; De Hevia et al., 2006) and also considered in the currently most influential model in numerical cognition research [i.e., the Triple Code Model (Dehaene, 1992; Dehaene and Cohen, 1997; Dehaene et al., 2003)]. Behavioral (e.g., Dehaene et al., 1993; Fischer, 2001, 2003) as well as neuropsychological (e.g., Zorzi et al., 2002) data provide evidence for an automatic activation of number magnitude on an analogous left-to-right oriented number line in Western cultures (see Shaki et al., 2009, for other cultures). Against this background, it is interesting to take a closer look at the development of the mental number line representation in children.

A standard task to make inferences about the development of the mental number line is the number line estimation task (e.g., Siegler and Opfer, 2003; Geary et al., 2008; Whyte and Bull, 2008; see also Petitto, 1990) also known as number-to-position task (e.g., Berteletti et al., 2010). In this task participants are required to estimate the spatial position of a given number on an empty number line with labeled endpoints defining the numerical range covered (e.g., 0–100; e.g., Siegler and Opfer, 2003). Usually, the deviance of the estimated position of a number from its correct position is interpreted to provide information about how the mapping of numbers to space and thus the mental number line representation develops. Generally, estimation performance is more error-prone in younger children as they tend to overestimate the position of relatively small numbers to the right (i.e., placing 9 at about the position of 40; Moeller et al., 2009). As a consequence, the positions of relatively large numbers are compressed toward the end of the scale which results in relatively high estimation errors (Siegler and Opfer, 2003; Booth and Siegler, 2006; Laski and Siegler, 2007). To account for this estimation pattern, Siegler and colleagues proposed children's estimations to represent a quite isomorphic reflection of a logarithmic underlying representation of number magnitude as the authors found a logarithmic function to fit the observed estimation pattern best. With increasing age and experience, however, the authors suppose children to develop a linear representation of number magnitude reflected by an estimation pattern fitted best by a linear function. This representational change, also referred to as log-to-linear shift, is interpreted to reflect the development toward a linear representation of number magnitude in older children and adults (Siegler and Opfer, 2003; Siegler and Booth, 2004; Booth and Siegler, 2006, 2008).

However, the conclusion of such a representational shift as drawn by Siegler and colleagues is currently discussed controversially with respect to both theoretical but also methodological issues (e.g., Barth and Paladino, 2011; Barth et al., 2011; Moeller and Nuerk, 2011; Slusser et al., 2013; see also Ebersbach et al., 2013 for an overview). From a theoretical point of view, there are alternative accounts to explain the developmental changes in number line performance. A seemingly logarithmic response pattern may be accounted for by a two- or multi-linear fitting procedure, while a seemingly linear pattern may be accounted for by proportion judgment. As regards logarithmic fitting, Moeller et al. (2009; see also Helmreich et al., 2011) observed that a two-linear model suggesting separate but linear representations for one- and two-digit numbers predicts the estimation performance of first-graders in a 0–100 number line task even better than a log-arithmic model. Theoretically speaking, the results of Moeller and colleagues do not indicate children's estimation pattern to directly reflect their spatial representation of number magnitude. Rather, they emphasize the importance of understanding the place-value structure of the Arabic number system: With increasing age and experience children master the integration of tens and units into the base-10 place-value structure of the Arabic number system and the separate representations are then integrated to result in a linear estimation pattern (Moeller et al., 2009; Helmreich et al., 2011; see also Ebersbach et al., 2008, for a similar two-linear approach). Another argument challenging the hypothesis of a representational log-to-linear shift was suggested by Barth and Paladino (2011) addressing seemingly linear fittings (see also Slusser et al., 2013). These authors suggested the standard number line estimation task to be more of a proportion judgment than a number magnitude estimation task. Barth and her colleagues argue that the to-be-estimated numbers are not considered in isolation but always in relation to reference points such as the start and endpoint of the given number line or its half. Their claim is methodologically corroborated by fitting results for power models usually used in proportion-judgment context (e.g., Spence, 1990) which provided the best fit for children's estimation performance: In contrast to a linear model which cannot account for systematic biases at reference points the authors found that either one- (considering start and endpoint as references; cf. Spence, 1990) or two-cycle power models (i.e., considering start and endpoint as well as their mean as reference points; cf. Hollands and Dyre, 2000) fitted 7-year-old children's estimation patterns on a 0–100 scale best (Barth and Paladino, 2011). From a theoretical point of view, Barth and colleagues suppose the standard (bounded) number line estimation task to reflect the application of proportion-judgment strategies rather than providing a direct measure of the spatial representation of number magnitude. This is corroborated by the finding that with increasing age more reference points are considered for estimation performance (Slusser et al., 2013).

The argument that the traditional number line estimation task induces strategies of proportion judgment was further corroborated by Cohen and Blanc-Goldhammer (2011). They observed smaller standard deviations of adults' estimations close to reference points but larger standard deviations between these points resulting in a characteristic *M*-shaped distribution. The validity of

such a pattern to indicate the use of reference points was also corroborated when evaluating eye fixation data (see Schneider et al., 2008, for children's eye fixation data; see also Sullivan et al., 2011, for adult data).

Considering a recently introduced new version of the number line estimation task (see below for more details) one aim of the current study was to evaluate whether proportion-judgment strategies found in bounded number line estimation are a generalizable characteristic of number line estimation and how the application of this strategy is related to age.

Despite the debate on the nature of the numerical representations and processes underlying number line estimation performance there is accumulating evidence suggesting that number line estimation performance is not only systematically related to actual numerical performance but also predictive of future numerical development. For instance, the acuity of children's mental number line representation as assessed by the linearity of children's number line estimations was found to be positively correlated with other numerical competencies such as numerosity estimation or numerical magnitude comparison (Booth and Siegler, 2006; Laski and Siegler, 2007) but also more complex arithmetic indices such as actual addition performance (Booth and Siegler, 2008). In the same study, children's number line estimation performance was also a reliable predictor of the ability to learn new addition problems (Booth and Siegler, 2008; see also Gunderson et al., 2012; Muldoon et al., 2013, for longitudinal evaluations of the relationship between number line estimation and children's mathematical development). Finally, there is now even first evidence from intervention studies suggesting a causal relationship between the acuity of the mental number line representation and more complex numerical/arithmetic abilities. For instance, Siegler and Ramani (2009; see also Ramani and Siegler, 2011) observed that playing simple linear number board games not only improved children's number line estimation performance significantly but also that this training effect generalized to their arithmetic competency (see also Fischer et al., 2011, for the validity of embodied experiences of spatial number magnitude; Kucian et al., 2011, for similar evidence in children with dyscalculia).

Taken together, it can be noted that number line estimation performance is a reliable predictor of actual and future numerical competencies even though it is still under debate what exactly is assessed by the number line estimation task in its standard bounded version with given start- and endpoint.

Cohen and Blanc-Goldhammer (2011; see also Booth and Siegler, 2006 for a somewhat similar task) proposed a new unbounded version of the number line estimation task without a predefined fixed endpoint. Instead, a unit (i.e., the distance between 0 and 1) is given together with a start point allowing for the estimation of the spatial position of a presented target number on a number line. Importantly, evaluation of participants' estimation pattern corroborated their hypothesis that this task version provided a less biased measure of the mental number line representation: There were no indications of systematic biases reflecting the use of reference points. Moreover, variability of participants' estimation errors increased continuously with number magnitude. This is in line with the assumption of a linear

mental number line representation with scalar variance (Gibbon, 1977; Gibbon and Church, 1981; Whalen et al., 1999). This scalar variance hypothesis suggests that the spacing between adjacent numbers on the mental number line is equidistant while representational uncertainty increases with the magnitude of the numbers. Against this background, the authors concluded that the unbounded number line estimation task seems to provide a more pure measure of the underlying mental number line representation as compared to the traditional bounded version of the task.

The only published data on the unbounded number line estimation task are from adult participants, however, number line estimation tasks are used much more prominently in the assessment of children's mental number line representation. Therefore, the objectives of the current study were straightforward.

We wished to evaluate how far the results of Cohen and Blanc-Goldhammer (2011) generalize to children's estimation performance. Therefore, we recruited a broad sample of primary school children from grade one through four as well as a sample of adult controls to perform both tasks, the new unbounded and the standard bounded version of the number line estimation task. Because there are no data available on children performing the unbounded number line estimation task, hypotheses were derived from recent data for the bounded number line estimation task. Slusser et al. (2013) observed children at the age of 7–8 to make use of 2 or 3 reference points (start-, end- and midpoint) to increase their estimation accuracy. In contrast, for younger children's estimation patterns indications for such a proportion-judgment strategy were less obvious. For five-year-olds the authors even suggested that children might have ignored the endpoint of the scale treating the task as an 'open-ended magnitude judgment' (Slusser et al., 2013, p. 203) comparable to an unbounded version of the number line estimation task. Furthermore, in an eye-tracking study Schneider et al. (2008) corroborated the assumption of qualitative differences between estimation strategies between relatively younger and older children. In particular, they observed that third-graders targeted their eye fixations more directly toward to-be-expected reference points (i.e., start, middle, and endpoint of the number line) than did younger children. Against this background, we expected a change in estimation strategies in the bounded number line task to occur from grade three at the latest with more pronounced indications for the use of reference points for older children. In contrast, based on the results for the unbounded number line estimation task in adults (Cohen and Blanc-Goldhammer, 2011) indicating a less biased measure of number line estimation no such qualitative change of estimation strategy was expected for the unbounded number line estimation task.

To pursue these hypotheses we evaluated two different aspects of our participants' estimation performance in accordance with the proceeding of Cohen and Blanc-Goldhammer (2011). First, we appraised participants' estimation patterns by fitting different kinds of models. Additionally, we considered the distribution of participants' estimation errors. Taken together, these two aspects should answer the questions whether there are qualitative differences in solution strategies (number line estimation vs. proportion judgment) between (i) the bounded and the unbounded version of the number line estimation task in children and (ii) at what age such differences emerge.

METHODS

PARTICIPANTS

A cross-sectional sample of 233 primary school children [65 first-graders (31 girls; mean age: 6;7 years, $SD = 3.90$ months), 61 second-graders (32 girls; mean age: 7;7 years, $SD = 5.58$ months), 59 third-graders (23 girls; mean age: 8;8 years, $SD = 6.35$ months) and 48 fourth-graders (27 girls; 9;8 years, $SD = 5.91$ months)] was assessed on a battery of basic numerical tasks including number line estimation to investigate the development of numerical competencies. Children were tested three months after class started. All children participated voluntarily and were included in the sample only after their parents provided a signed informed consent form. In addition, a control sample of 68 university students (56 females; mean age: 23;5 years, $SD = 4;8$ years) volunteered to perform the number line tasks.

STIMULI

For the different age groups we used different number scales covering the ranges that are taught in the respective grades and can thus be considered more or less familiar to the children tested, this means that children should possibly be able to infer the midpoint of the respective range (first-graders: 0–10; second-graders: 0–20; third-graders: 0–100; fourth-graders: 0–1000; adults: 0–10,000). At all age groups 20 stimuli for the bounded number line task were chosen to allow for reliable identification of possible proportion-judgment strategies, resulting in more items at the suggested reference points (cf. Barth et al., 2011). A total of four items was displayed on one DIN A4 sheet with the start-point of the number lines being varied horizontally to prevent participants from relying on estimates of previous trials as possible anchor points. All number lines were 20 cm long with labeled endpoints below and the to-be-estimated number placed above the middle of the number line (see **Figure 1** for a schematic illustration). In all number ranges two practice items (exception: number range 0–10 with only one practice item) ensured participants understanding of the task and were shown on the first page prior to the critical trials. Different from the bounded number line estimation task we did not change the range covered by the unbounded number line task for the different age groups. The same physical length of 20 cm was used for the unbounded and the bounded task to enhance comparability between task versions. The unit indicating the distance between 0 and 1 was depicted below the start-point. The to-be-estimated numbers were presented above the start point (see **Figure 1** for a schematic illustration). The numerical length of the unbounded number line was 29. Only items in the range from 0 to 20 were used to assess unbounded number line performance, leaving enough space between the largest numbers and the physical endpoint. A total of 15 items were presented as numbers 0, 1, 5, 11, and 20 were excluded and 10 served as practice item. Again, four items were presented on one DIN A4 sheet arranged in the same way as the bounded task items.

FIGURE 1 | Schematic illustration of the bounded number line task and the new unbounded version of the task.
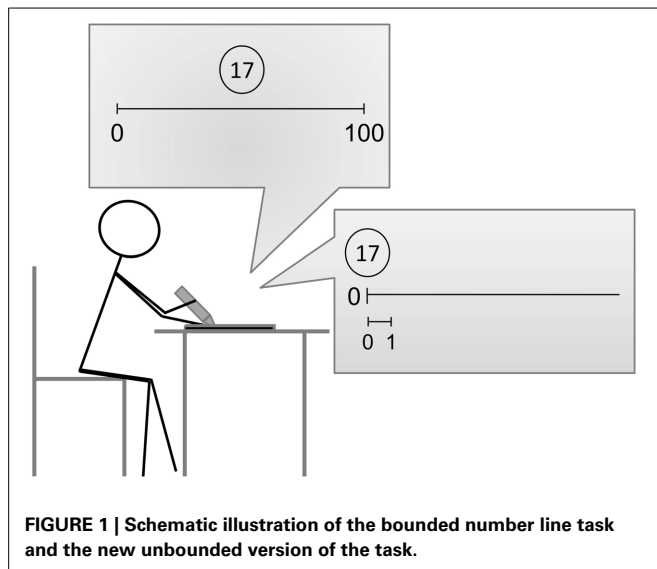
PROCEDURE

All tasks were administered in group settings. For the bounded number line task, children received the oral instruction that they are presented with a special number line which only has a start- and endpoint but no further numbers in between. Then the task was explained (in German) as follows: "Look at the number printed above the number line—where do you think this number goes between 0 and X. Please mark your estimate on the line." Importantly, no numbers apart from the start and the endpoint were indicated. In the unbounded task, the instruction was similar: Children were told that there is no end to the number line but that they can see how long the distance from 0 to 1 is. Further instructions were adapted to the task: "Look at the number printed above the number line—where do you think this number goes?" Again, no other numbers were indicated to participants. Adults were provided with written instructions. Neither age group received feedback as to the correctness of any of the items. Task order was the same for all age groups: participants started with the bounded followed by the unbounded number line estimation task.

ANALYSES

As variables of interest we evaluated children's mean estimates (indicating their estimation performance) as well as the standard deviation of children's percent absolute error (PAE = |Estimate − Target number|/ Scale *100; cf. Siegler and Booth, 2004, reflecting the variability of their estimates). In line with the procedure of Ashcraft and Moore (2012) we ran a contour analysis contrasting the variability of children's estimation errors at and in between possible reference points (using *t*-tests) for both the bounded and unbounded estimation task separated for each age group. Therefore, standard deviations of the PAEs of the two target numbers closest to the origin, the first quartile, the midpoint, the third quartile and the endpoint, were pooled. In case target number and reference point were identical (e.g., item 15 in the unbounded task conforms to the third quartile) the item itself plus the two

closest target numbers were considered. For first-graders' on the 0–10 bounded number line task only one item was considered for the origin and endpoint, respectively, as there were only 8 target numbers.

In addition, we evaluated the goodness of fit of several models used in previous studies to mathematically reflect children's estimation performance (e.g., linear, power models, etc.). To fit models Matlab 7.14 was used applying the trust region algorithm for the fitting of non-linear models. For both number line tasks we estimated the fit of the same models thereby differentiating grossly between three families of functions corresponding to different estimation strategies: (i) direct estimation strategies should be indicated by the superior fit of the linear and (unbounded) power function (cf., Slusser et al., 2013), (ii) proportion-judgment based estimation strategies should be indexed by one- and two-cycle power models (cf. Barth et al., 2011), and (iii) dead-reckoning strategies should be reflected by dual and multi scallop models (for scallop models see Cohen and Blanc-Goldhammer, 2011).

Linear models were fitted with two free parameters (i.e., the intercept and the slope). The unbounded power model had one free parameter (i.e., the exponent) while dual scallop and multi scallop models were fitted with two free parameters (i.e., the exponent and the size of the working window). The linear and the unbounded model allow for identifying direct estimation strategies, with no application of an additional strategy like dead-reckoning or proportion judgment. In contrast, cyclic power models suggest that participants use at least two references point (start and end point for the one-cycle model whereas the two-cycle model indicates the use of an additional central reference point). Cyclic models were fitted with one free parameter (i.e., the exponent determining the shape of the power function). Dual and multi scallop models are well suited to find out whether participants applied a dead-reckoning strategy. Thereby, participants first estimate a particular working window of numbers (e.g., 5) and then use multiplies of this working window to estimate the position of higher numbers. The dual scallop model allows for identifying participants, who applied the working window twice and the multi scallop model participants, who repeated their working window multiple times. As identification of such dead-reckoning strategies was not at the heart of this study, we summarized results of the respective models in the category of "others."

However, different from testing the scallop models in the bounded condition applying cyclic power models in the unbounded task is not as straightforward as it seems to be at first glance. Importantly, cyclic power models require definition of an upper bound, which can be easily specified in a bounded number line task. However, for the unbounded number line task such an upper bound does not exist *per se*. Theoretically, participants might have used the end of the physical line as an upper bound or the largest number which they had to estimate. Since these strategies might vary between participants, a fixed upper bound for testing cyclic models in the unbounded task cannot be used. Therefore, this upper bound has to be estimated by the fitting procedure (cf. Barth et al., 2011). The range of the parameter accounting for the upper bound was allowed to vary between 19

and 29, corresponding to the largest target number (19) and the numerical end of the line (29).

Models were compared by calculating AICc (Akaike information criterion with a correction for finite sample sizes) values for each participant (e.g., Burnham et al., 2011; see also Cohen and Blanc-Goldhammer, 2011, for a similar procedure). Lower AICc values were then interpreted as superior fit of either model[1].

RESULTS

In total, 61 participants (18 first-graders, 18 second-graders, five third-graders, five fourth-graders and 15 adults) were excluded from final analyses as they had missing data on at least three items within one of the tasks and/or showed an estimation pattern indicating insufficient understanding of the task (e.g., marking the middle of the number line for all trials). Furthermore, individual estimates that differed more than ± 3 SD from the age groups' mean estimate were also excluded. It is important to note that this trimming procedure did not change results substantially.

ESTIMATION PATTERNS AND MODEL FITTINGS

Mean estimates were calculated separately for all age groups and plotted as a function of target number to look for obvious indications of the use of reference points (see **Figure 2**). We found that for all age groups mean estimates increased steadily with increasing size of the target number independently of task version. Only first graders' bounded number line estimates obviously differed in distribution from older children's and adults' bounded estimates (see **Figure 2A**, left column). In general, first graders seemed to underestimate larger numbers as they did not produce estimates larger than about 6 on the 0–10 bounded number scale.

For the unbounded number line task the distributions of estimates look very similar for the different age groups (see **Figures 2A–E**, right column). It is notable, that first-graders and adults overestimated numbers toward the end of the scale (i.e., numbers close to 20; see **Figure 2A**, right column), seemingly they often tended to locate larger numbers toward the end of the physical line (which was at 29).

---

[1] Please note that we previously advocated bi-linear or multi-linear fittings of children's number line estimations whereas we fitted one-cycle and two-cycle models (cf. Barth and Paladino, 2011; Slusser et al., 2013) in the current study. Importantly, we wish to emphasize that this is not a contradiction. In recent studies we fitted seemingly logarithmic estimation patterns using a bi-linear model and argued that what seems logarithmic actually reflects separate representations for single- and two-digit numbers possibly indicating insufficient place-value understanding (Moeller et al., 2009; Helmreich et al., 2011; Moeller and Nuerk, 2011). However, proper understanding of the base-10 place-value structure is actually a mandatory prerequisite for any proportional strategy because the proportions applied must be represented correctly (or at least roughly) to produce correct estimates. The idea that successful base-10 place-value understanding is inevitably necessary for the task and that proportional strategies are employed in later developmental stages are not incompatible with each other: The multi-linear account represents an alternative for seemingly logarithmic estimation patterns whereas the proportion-judgment account reflects an alternative account for seemingly linear (as compared to logarithmic) estimation patterns. Because children of all age groups consistently showed seemingly linear rather than logarithmic estimation patterns in the current study bilinear accounts, which suppose insufficient base-10 place-value understanding, are not applicable here.
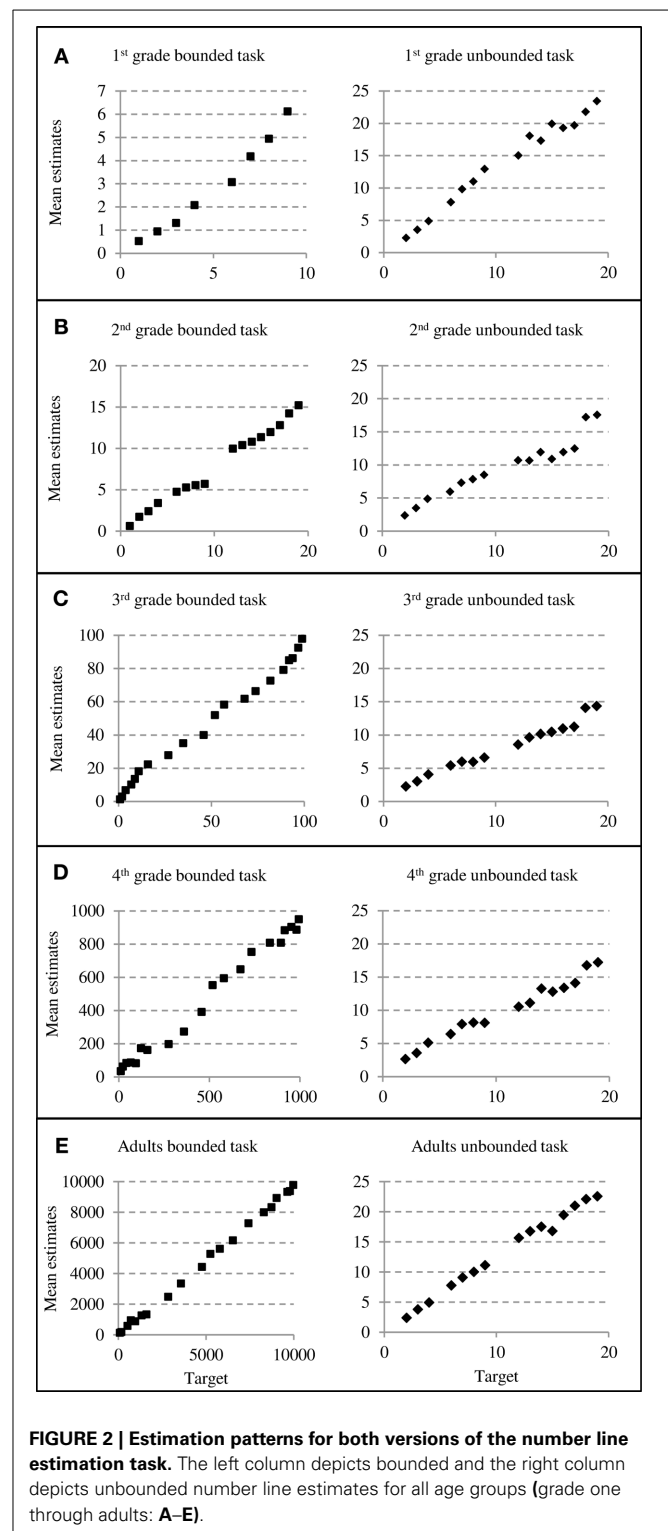
FIGURE 2 | Estimation patterns for both versions of the number line estimation task. The left column depicts bounded and the right column depicts unbounded number line estimates for all age groups (grade one through adults: **A–E**).

Because plotting estimates as a function of target number only allows for visual inspection to conclude whether children did or did not use proportion-based estimation strategies, we fitted estimates with different types of models. **Table 1** depicts frequencies of best fitting models for the different age groups separated

Table 1 | Absolute and relative frequency (percentages) of best fitting models indicating direct estimation, proportion judgments or other estimation strategies (left column) and detailed distribution of best fitting model of participants' estimates (right column) separated for bounded and unbounded number line tasks and age groups.

	Estimation strategies			Model fittings					
	Direct	Proportional	other	Linear	Unbounded power	Dual scallop	Multi scallop	One- cycle	Two- cycle
BOUNDED TASK									
1st grade	**40 (85)**	7 (15)	0	**30 (64)**	10 (21)	0	0	4 (9)	3 (6)
2nd grade	**32 (74)**	9 (21)	2 (5)	**22 (51)**	10 (23)	1 (2)	1 (2)	5 (12)	4 (9)
3rd grade	17 (31)	**37 (69)**	0	4 (7)	13 (24)	0	0	**25 (46)**	12 (22)
4th grade	15 (35)	**28 (65)**	0	5 (12)	10 (23)	0	0	8 (19)	**20 (47)**
Adults	**37 (70)**	16 (30)	0	12 (23)	**25 (47)**	0	0	7 (13)	9 (17)
UNBOUNDED TASK									
1st grade	**40 (85)**	4 (9)	3 (6)	15 (32)	**25 (53)**	2 (4)	1 (2)	4 (9)	0
2nd grade	**39 (91)**	4 (9)	0	17 (40)	**22 (51)**	0	0	3 (7)	1 (2)
3rd grade	**50 (93)**	0	4 (7)	12 (22)	**38 (70)**	2 (4)	2 (4)	0	0
4th grade	**39 (91)**	2 (5)	2 (5)	10 (23)	**29 (67)**	1 (2)	1 (2)	2 (5)	0
Adults	**45 (85)**	2 (4)	6 (11)	16 (30)	**29 (55)**	2 (4)	4 (8)	2 (4)	0

The best fitting models are indicated in bold script.

for both, task and strategy applied (parentheses show relative frequencies).

First- and second graders seemed to use a direct estimation strategy solving the bounded number line estimation task as the linear model provided the best fit (for 64% of the first- and 51% of the second-graders). Interestingly, this pattern changed for third- and fourth-graders: Only 31% of the third-graders' and 35% of the fourth-graders' estimates were accounted for best by models indicating direct estimation strategies (i.e., linear and unbounded power models). Instead, one- or two-cycle models provided a better fit, clearly indicating the use of reference points and thus proportion-judgment strategies (Barth and Paladino, 2011). In detail, 68% of third-graders' estimates were fitted best by cyclic power models: 46% by one-cycle and 22% by two-cycle models. Thus, most third-graders seemed to consider two reference points (i.e., start- and endpoint). Moreover, 66% of fourth-graders' estimates were also accounted for best by cyclic power models. The high percentage for two-cycle models (47%) indicated the prominent use of three reference points.

Unexpectedly, a direct estimation strategy was also observed for the majority of adult's estimates (70%) as the single scallop model provided the best fit for 47% of participants' estimates. At first glance, this result pattern seems to contradict our hypothesis and also previous results of Cohen and Blanc-Goldhammer (2011). However, a closer look at the estimation pattern clarified this: as adults show estimation patterns with very small PAEs all model fittings were more or less identical as indicated by the respective adjusted R^2 (linear model: *mean adj.R^2 = 0.985*, unbounded power model: *mean adj.R^2 = 0.984*, dual scallop model: *mean adj.R^2 = 0.983*, multi scallop model: *mean adj.R^2 = 0.983*, one-cycle model: *mean adj.R^2 = 0.981*, two-cycle model: *mean adj.R^2 = 0.981*). Thus, as power models with an exponent of 1 are basically similar to a linear function without an intercept, selection of best fitting model does not provide sufficient evidence to reliably differentiate between estimation strategies. Yet, a

closer inspection of (adults') estimation errors is informative (see below).

Regarding model fittings for estimates in the unbounded version, results are more consistent: Independent of age group, the majority of participants was always classified to use direct estimation strategies as an unbounded power model provided the best fit for their estimates. According to Cohen and Blanc-Goldhammer (2011), this model indicates that participants directly estimated targets' locations (see also Slusser et al., 2013). Dual and multi scallop models fitted best for only a few participants' estimates, most frequently adults, as did cyclic models. Taken together, these data do not corroborate the notion of a prominent use of specific strategies such as proportion-judgment or dead-reckoning in unbounded number line estimation.
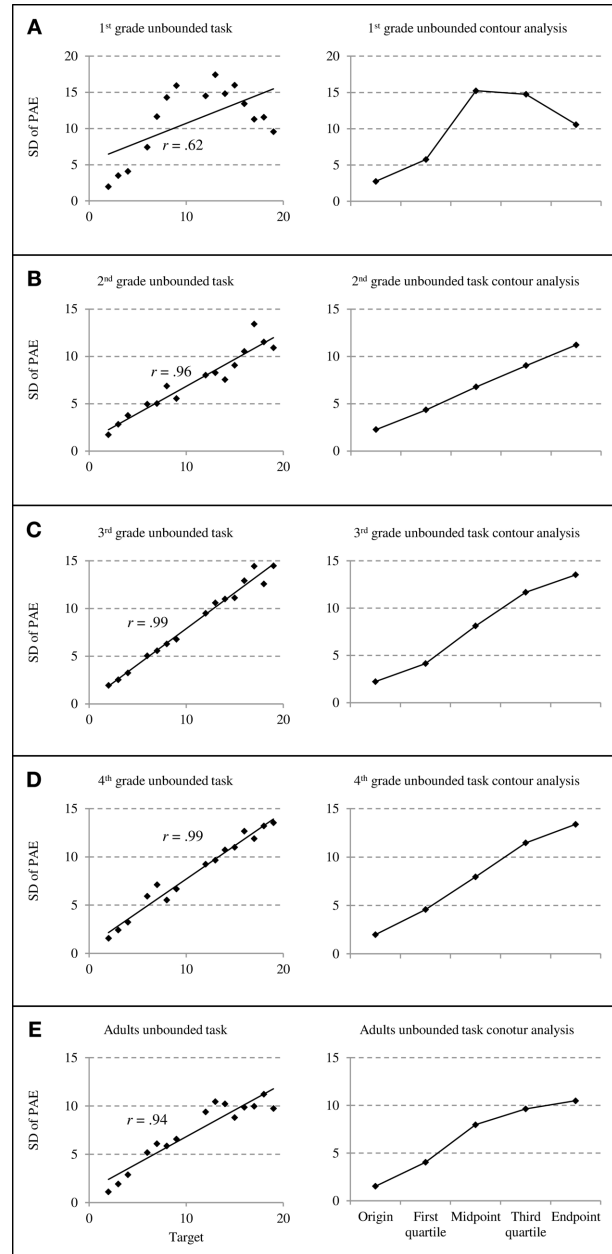
PAE DISTRIBUTION

On the panels of **Figures 3** and **4**, we plotted the mean standard deviations of PAE as a function of target number (see left column) and the standard deviations of PAEs at targets close to specific reference points (i.e., origin, midpoint, and endpoint) and in between reference points (first quartile, third quartile) to be compared in a contour analysis (right column; cf. Ashcraft and Moore, 2012).

Even from visual inspection of bounded number line estimation performance, it is obvious that between grades two and three (**Figures 3B,C**) a change in children's estimation strategies seems to occur. First- and second-graders' PAE variability increased significantly with increasing target numbers as indicated by both, the more detailed distribution of PAE variability (**Figures 3A,B**, left column) as well as the contour analyses (**Figures 3A,B**, right column). Correlating SD of PAE and size of target number revealed significant correlations of $r = 0.99$ ($p < 0.01$) for both age groups. In contrast, children from grade three on showed *M*-shaped patterns of PAE distribution and no significant correlation between SD of PAE and size of target
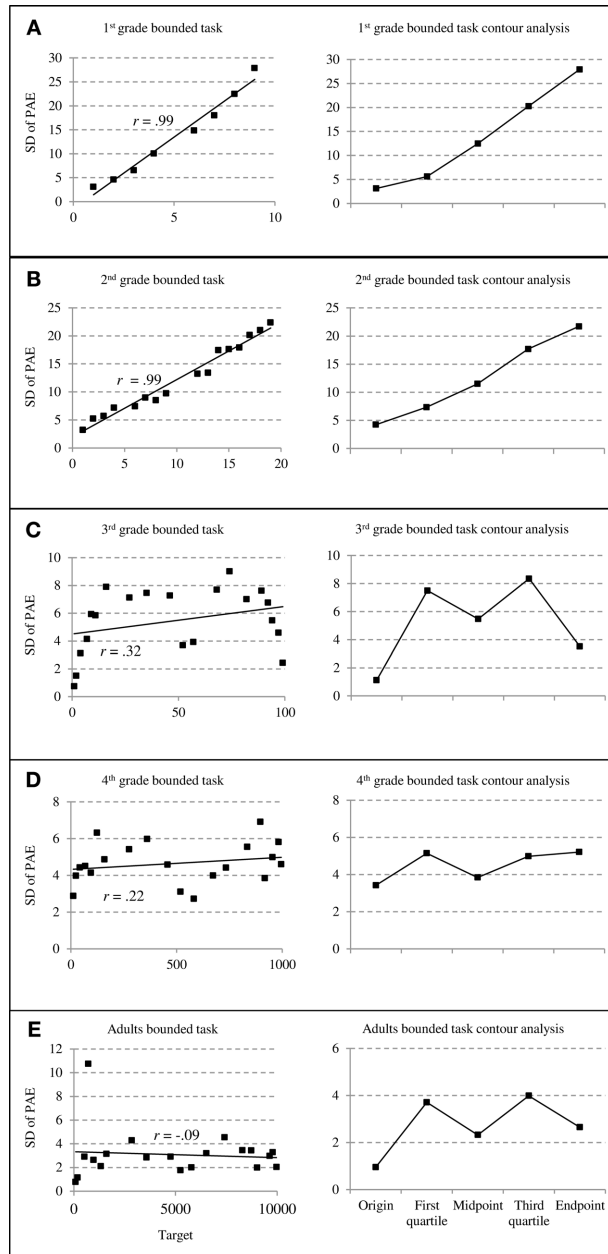
FIGURE 3 | Standard deviations of percent absolute errors (PAE) for bounded number line estimation performance. The left column depicts the relationship between standard deviations of PAEs and target numbers whereas the right column depicts results of counter analyses summarizing standard deviations of PAEs at specific reference points (cf. Ashcraft and Moore, 2012; grade 1 through adults: **A–E)**.

FIGURE 4 | Standard deviations of percent absolute errors (PAE) for unbounded number line estimation performance. The left column depicts the relationship between standard deviations of PAEs and target numbers whereas the right column depicts results of counter analyses summarizing standard deviations of PAEs at specific reference points (cf. Ashcraft and Moore, 2012; grade 1 through adults: **A–E)**.

number (all $r < 0.32$, all $p > 0.18$). This means that children's estimations varied less at and around the to-be-expected reference points (i.e., start and endpoint as well as the midpoint of the scale) whereas PAE variability was reliably larger in between these reference points. Additionally, the same patterns were also present for fourth-graders and adults, especially when taking a closer look at the distribution of PAE variability plotted as

target function (see **Figures 3C–E**, left column). Importantly, these (indicated) *M*-shaped patterns of error distribution are characteristic for proportion-judgment strategies (cf. Cohen and Blanc-Goldhammer, 2011). Generally, statistical evaluation by the contour analyses (see **Figure 3**, right column) substantiated these *M*-shaped patterns. The *t*-tests to evaluate whether PAE variability is indeed reduced at/around suspected reference points (start-,

mid and endpoint) compared to in between them (first and third quartile) revealed no significant differences for first- and second-graders' PAE variability (both $t < 0.88$, both $p > 0.40$, one-sided) but indicated (marginally) significant smaller PAE variability at/around reference points for third-graders, adults (both $t > 4.10$, both $p < 0.01$, one-sided), and fourth-graders [$t(8) = 2.19$, $p = 0.06$, one-sided].

In contrast to this, for the unbounded number line task similar PAE distributions were observed across all age groups: PAE variability increased monotonously with target number (see **Figures 4A–E**) resulting in significant correlations between SD of PAE and size of target number for all age groups (from $r = 0.62$ to $r = 0.99$, all $p < 0.05$). This pattern of linearly increasing error variability was most prevalent for second-, third-, and fourth graders. However, for first-graders we observed PAE variability to decrease toward the end of the scale (see **Figure 4A**) while PAE variability remained constant for adults' estimates of larger target numbers (see **Figure 4E**). This pattern was due to the fact that first-graders and adults placed larger numbers toward the end of the physical line, thereby increasing PAEs but reducing their variability or holding it constant, respectively. Importantly, the t-tests statistically evaluating the contour analysis did not reveal any significant differences in PAE variability at/around reference points compared to PAE variability in between these reference points (all $t < 0.50$, all $p > 0.63$).

DISCUSSION

The current study set off to investigate the development of children's spatial representation of number magnitude by comparing their estimation performance in both a standard bounded as well as a new unbounded version of the number line estimation task. Such a direct contrast of the two versions of the task for children is of particular interest as a recent study with adult participants (Cohen and Blanc-Goldhammer, 2011) indicated that the standard, bounded number line estimation task seems to induce proportion-judgment strategies whereas the new, unbounded version of the task was supposed to provide a more pure measure of the underlying spatial magnitude representation. This was concluded by Cohen and Blanc-Goldhammer (2011) from the fact that no reduction of error variability at specific reference points was observed in unbounded number line estimation indicating that this task is better suited to make inferences on the representation of integer numbers along the mental number line. To investigate the development of possible differences between these two versions of the task we assessed children from first to fourth grade on the bounded as well as the unbounded number line estimation task in a cross-sectional design. In line with recent data (Schneider et al., 2008; Slusser et al., 2013) we expected to observe evidence for proportion-judgment strategies in children from grade three at the latest for the bounded but not the unbounded number line estimation task. The present data partially corroborated this hypothesis as we observed a qualitative change in estimation performance for bounded but not unbounded number line estimation with increasing age, in particular between second and third grade. Note, however, that Slusser et al. (2013) observed evidence for the predominant use of proportion-judgment strategies already in seven-year olds whereas in the current study this

was only the case for third-graders and older children. Cultural or school characteristics or task attributes may explain this slight difference.

In the following, we will first discuss the differential development in the bounded and unbounded version of the number line estimation task before elaborating on the broader implication for research on children's numerical development.

DIFFERENTIAL DEVELOPMENT OF NUMBER LINE ESTIMATIONS IN BOUNDED vs. UNBOUNDED NUMBER LINE TASKS

First- and second-graders' bounded and unbounded number line estimations indicated no substantial differences between the estimation patterns and error distributions for the two versions of the number line estimation task in our study. This corroborates our hypothesis that indices for the use of proportion-judgment strategies (considering at least two reference points) may only occur after a certain level of proficiency has been reached (see also Slusser et al., 2013 for a similar argument, however, for an earlier start of proportion judgments). Importantly, this interpretation is corroborated by the results of the model fittings: For the bounded estimation task, we found the estimation pattern of most children to be fitted best by linear functions instead of cyclic models not indicating the use of proportion-judgment strategies. In line with this, unbounded power models were observed to fit best the estimates of most children in the unbounded number line estimation task—again not indicating the use of proportion-judgment strategies. Instead the prominently observed models indicate that children directly estimated target numbers without using reference points other than the start point (cf. Cohen and Blanc-Goldhammer, 2011).

Apart from this general pattern there was an interesting finding for the first-graders in our sample. We observed that first-graders overestimated positions of large numbers in the unbounded number line task. This means that they placed numbers close to 20 (which was the maximum of the number range assessed) even beyond than necessary (i.e., further to the right, see **Figure 2**, Panel A). Because we also observed a decrease of the variation of the estimation errors toward the end of the scale, this probably indicates that first-graders used the length of the physical line as any kind of orientation. However, because there was no numerical endpoint indicated and there is evidence that relatively younger children tend to even ignore the upper bound when given (at least they do not seem to use it systematically as a reference point, Slusser et al., 2013) we are confident that this does not suggest the use of proportion-judgment strategies. This is also backed by the modeling results with no indications for cyclic models to fit the data best. Rather, children seemed to consistently overestimate the target numbers with the largest numbers seeming so large to them, that they locate them toward the end of almost any unbounded number line. Synced with the fact that even first graders are usually able to adhere to the ordinal sequence of the numbers in number line estimations (e.g., Moeller et al., 2009) it is just a consequence of such behavior that error variation decreases toward the end of the physical line.

In contrast to first- and second-graders and in line with previous studies investigating bounded number line estimation (e.g., Slusser et al., 2013) estimation performance of relatively older

children revealed explicit differences between the bounded and unbounded version of the number line task. For the unbounded number line estimation task estimation patterns as well as the monotonously increasing variation of estimation errors did not indicate the use of reference points. Again this was corroborated by the modeling results as we found the estimation patterns of the vast majority of children (third- and fourth-graders) to be fitted best by models indicating direct estimation strategies. However, inspection of both, estimation patterns as well as error variability indicated that this did not hold for estimation performance in the bounded number line estimation task. Although estimation patterns looked rather linear one- and two-cycle power models provided the best fit for the majority of children's' estimates - clearly indicating the use of either two (i.e., start and endpoint) or three reference points (i.e., start, middle and endpoint; cf. Cohen and Blanc-Goldhammer, 2011; Slusser et al., 2013). Furthermore, the variation of estimation errors showed the characteristic *M*-shaped distribution indicating that error variability decreased at these respective reference points (cf. Cohen and Blanc-Goldhammer, 2011). Thus, our results are in line with recent evidence suggesting that relatively older children probably from grade three on (Schneider et al., 2008, but see Slusser et al., 2013 for proportional judgment from grade two on), systematically rely on proportion-judgment strategies in number line estimation - but only so when performing the bounded number line task.

A similar pattern was observed when looking at adults' estimation performance. Although model fittings indicated adults to use direct estimation strategies a closer inspection of PAE variability was also informative. The *M*-shaped distribution of PAE variability clearly indicated the use of proportion-judgment strategies. Synced with the fact that model fittings can hardly differentiate between very accurate and basically linear estimation patterns the fitting results for adult participants should be considered only cautiously. In contrast, adults' estimates on the unbounded number line revealed no indication for the systematic use of proportion-judgment strategies. However, different from second- to fourth-graders and comparable to first-graders, adults seemed to use the end of the number line as some kind of endpoint. Not only did adults overestimate numbers close to the largest target number assessed but their PAE variability remained constant for these target numbers as well. This result pattern, however, is in line with previous findings of Cohen and Blanc-Goldhammer (2011) who removed participants' responses for the largest items from further analyses because "the computer screen boundary acted as an artificial endpoint and skewed these data low" (Cohen and Blanc-Goldhammer, 2011, p. 335). Importantly, even though adults may have tried to figure out the endpoint of the number line by locating the largest targets toward the end of the physical line, model fittings as well as visual inspection of PAE variability did not provide any evidence for the use of specific estimation strategies (i.e., prominent use of proportion-judgment or dead-reckoning strategy) in unbounded number line estimation.

Taken together and in line with previous studies (cf. Cohen and Blanc-Goldhammer, 2011; Slusser et al., 2013), our results suggest that the standard bounded number line estimation task seems to induce specific proportion-judgment strategies for

relatively older children and adults. Therefore, these data add to recent evidence challenging the view that the bounded version of the number line estimation task allows for direct inferences about the nature of the spatial representation of number magnitude (see also Barth and Paladino, 2011; Karolis et al., 2011). Cohen and Blanc-Goldhammer (2011) proposed that not the estimation pattern but the error variability found in number line estimation tasks allows inferences about the representation of the magnitude of numbers. In line with their data for adults we found the error variability to increase linearly in the unbounded number line estimation task indicating a linear number representation with scalar variance (e.g., Gibbon and Church, 1981; Brannon et al., 2001). However, most importantly following the rationale of Cohen and Blanc-Goldhammer (2011) our data suggest that the unbounded version seems to provide a purer measure of number line estimation performance in children as well—at least for relatively older children while we were not able to find systematic differences between performance in bounded and unbounded number line estimation for the relatively younger participants in our study (first-and second-graders). Thus, for relatively younger children both versions of the task may tap on number line estimation whereas for older children performance in the bounded version may be complemented by strategies other than number line estimation. This is of particular interest from a developmental point of view because performance in the bounded version of the number line estimation task has repeatedly been associated with actual as well as future numerical achievement (e.g., Booth and Siegler, 2008).

IMPLICATIONS FOR RESEARCH ON NUMERICAL DEVELOPMENT

Estimation performance in the bounded number line task is closely related to other numerical concepts, to some extent even causally (e.g., Booth and Siegler, 2008; Siegler and Ramani, 2009). Children with a more accurate linear representation are not only more proficient in other numerical tasks such as addition but are also better in learning new arithmetical problems. Yet, given the interpretation of Slusser et al. (2013), who suggested children to improve in number line estimation when able to consider more reference points to successfully apply proportion-judgment strategies, the question arises what it is that links performance in the bounded number line estimation task to other numerical concepts—acknowledging that it may not be (as originally proposed) the index of the underlying spatial magnitude representation? A possible account might consider the conceptual similarity of applying proportion-judgment strategies to some extent involve an understanding of part-whole relations and thus fractions and the concept of division (e.g., the midpoint requires an understanding of halving). Importantly, there is now accumulating evidence indicating that fraction understanding has a central role in numerical development as well as educational achievement, in particular beyond the first few years of schooling (see Siegler et al., 2013, for a review). Not only that high school students' fraction knowledge correlates very high with their actual mathematics achievement ($r > 0.80$). Fraction knowledge of relatively older children (i.e., fifth-graders) also predicts future algebra and overall mathematics achievement in high school even after taking into account covariates such as IQ, reading ability,

working memory, etc. (Siegler et al., 2012, see also Bailey et al., 2012; Booth and Newton, 2012 for the influence of fraction understanding on mathematics achievement). Importantly, this is in accordance with educational and instructional practice. Lack of fraction knowledge was ranked to be amongst the most important problems hindering students' algebra learning by a representative sample of 1000 US algebra teachers (National Mathematics Advisory Panel, 2008a). And as a consequence, the National Mathematics Advisory Panel (2008b) asserts that "the teaching of fractions must be acknowledged as critically important and improved before an increase in student achievement in algebra can be expected" (p. 18). In this vein, we propose that for older children the concept of proportionality also or predominantly drives the observed predictive power of estimation performance in the bounded number line estimation task for actual and future numerical and arithmetical achievement. In sum, it may not be the spatial representation of number magnitude also assessed in the unbounded number line task, but rather proportional strategies specific to the bounded number line task which are related to other arithmetic competencies.

LIMITATIONS AND PERSPECTIVES

Although it was not at the heart of the current study to compare the two versions of number line estimation tasks with respect to task difficulty, we wish to elaborate on potential concerns about the different number ranges assessed being responsible for the application of different strategies. The choice of different number ranges for the bounded task was based on the results of previous studies (cf. Slusser et al., 2013) which showed that already 8- to 10-year-olds are not only able to perform number line tasks in ranges up to 100,000 but also applied proportion-judgment strategies in these ranges. Therefore, the eventual concern that the bounded number line estimation task might have been more difficult for older children and adults simply because of the higher number ranges covered seems premature. This argument is further corroborated by a closer inspection of means and standard deviations of PAEs. As can be read from **Table 2** estimation errors were higher for bounded number line performance than for unbounded number line performance only for first- and second-graders (both $t > 3.1$, both $p < 0.01$) although the assessed number ranges in the bounded task were either the same or even smaller compared to the range assessed in the unbounded number line estimation task. Interestingly, the reversed pattern

was found for children older than grade two and adults: even though larger and supposedly more difficult number ranges were administered in the bounded task versions mean PAEs are higher for the unbounded task (all $t > 5.8$, all $p < 0.01$). This pattern nicely corroborates our hypothesis that older children and adults apply proportion-judgment strategies for solving bounded number line estimation (making these easier) whereas younger children are (not yet) able to apply such a strategy.

An additional argument corroborating this interpretation comes from the development of PAEs. Comparing mean PAEs within bounded number line estimation, an ANOVA revealed a main effect of age group [$F_{(4, 239)} = 68.75$, $p < 0.01$]. *Post-hoc* pairwise comparisons showed significant differences between mean PAEs of first-graders and second-graders with all other age groups (all $p < 0.01$). Third- and fourth-grader's as well as adult's mean PAEs for the bounded task did not differ significantly. For mean PAEs for unbounded estimation an ANOVA revealed a main effect of age group [$F_{(4, 239)} = 5.54$, $p < 0.01$]. *Post-hoc* tests indicated significantly higher PAEs for first-graders compared to second-graders and adults ($p < 0.05$, $p < 0.01$, respectively). Additionally, third-graders were indexed to differ significantly from adults regarding their mean PAE ($p < 0.05$). So, constant PAEs for third-, fourth-graders and adults in the bounded number line task indicate that the difficulty of the respective range assessed was approximately the same for the different age groups. Since differences in estimation errors of unbounded number line performance reveal no systematic pattern (see **Table 2**, notification) there are no obvious indications for strategy application.

However, besides these issues regarding the ranges used in bounded number line estimation other theoretical as well as methodological aspects which we did not control for explicitly might also play a role in number line estimation performance. For example, differences in the visual appearance of the tasks, or the missing second landmark in the unbounded task, in particular, might involve different processes of spatial recall. For instance it is interesting to note that Huttenlocher et al. (1994; see also Hund and Spencer, 2003; Schutte et al., 2011), observed that children's performance pattern in spatial recall tasks were very similar to estimation patterns found for bounded number line estimation. When supposed to remember the spatial location of a hidden toy, children also relied on the boundaries of sandboxes as location cues. While younger children (4- to 7-year olds) showed a memory bias toward the middle of the sandbox older children (10- to 11-year olds) showed a bias toward the first and third quartile of the box (speaking in terms of number line segmentation; Huttenlocher et al., 1994). Thus, the application of proportion-judgment strategies in bounded number line estimation might also be influenced by more general aspects of spatial cognition. This assumption is further corroborated when considering the influence of spatial attention for number line performance. LeFevre et al. (2010; see also LeFevre et al., 2013) observed that besides linguistic and quantitative processes, spatial attention is a unique precursor for early numeracy skills, also predicting number line estimation performance.

In sum, this correspondence asks for further studies to disentangle how number line estimation is influenced by both general

Table 2 | Mean PAE's (percent absolute error) and *SD* of PAEs for the range of the respective number line task separated for the different age groups.

Age group	Bounded task			Unbounded task		
	M	*SD*	Range	*M*	*SD*	Range
First grade	22.8	11.6	0–10	15.0	7.6	0–20
Second grade	14.7	9.2	0–20	10.7	4.4	0–20
Third grade	6.4	2.8	0–100	13.8	6.6	0–20
Fourth grade	6.0	2.1	0–1000	11.9	6.1	0–20
Adults	3.3	1.1	0–10,000	10.0	5.7	0–20

influences of spatial cognition and the particular spatial attributes of task presentation such as item placement or task instruction (cf. Barth and Paladino, 2011). It is well conceivable that such factors also influence estimation performance. For example, in our study, items of the bounded task were always depicted above the midpoint of the number line whereas items of the unbounded task were depicted above the start point. However, as our results obtained for adult participants are more or less identical to those of Cohen and Blanc-Goldhammer (2011) who controlled for item placement we would not assume children's estimation patterns to be distinct when item placement was controlled. In addition, there are other interesting and important research questions for future studies regarding the way performance in the unbounded number line estimation task relates to other actual but also future numerical and arithmetical competencies. Assuming unbounded number line estimation to provide a purer measure of the underlying spatial representation of number magnitude, evaluating its relationship with other numerical competencies might be of particular interest given the strong relationship observed for bounded number line estimation.

CONCLUSIONS

In the current study we directly compared children's estimations in the standard bounded as well as a new unbounded version of the number line estimation task. In line with recent research we found reliable evidence for the use of proportion-judgment strategies in bounded number line estimation for relatively older children (third- and fourth-graders) and adults adding to evidence suggesting that estimations in the bounded number line task are not reflecting an isomorphic measure of the mental number line of relatively older children and adults.

In contrast, there were no indications for the use of any strategies other than direct number line estimation for the unbounded number line estimation task which, thus, may be a valuable tool for assessing the spatial magnitude representation in a more unbiased way, at least for older children and adults. The fact that we did observe similar results for the bounded and the unbounded version of the task for first- and second-graders may indicate that both versions of the task might assess the same underlying representation for relatively younger children at least in number ranges familiar to the children assessed.

Taken together, the bounded and the unbounded number line estimation task seem to assess different representations and processes, although aiming to assess the same underlying spatial representation of number magnitude. Importantly, this has implications on a broader level. As estimation performance in the bounded number line task is not only correlated with but even causally related to other numerical and arithmetic competencies, future research is necessary to investigate whether it is indeed the spatial representation of number magnitude (assessed by the bounded number line task) or rather the concomitantly assessed proportion understanding which predicts future numerical competencies and achievement.

REFERENCES

Ashcraft, M. H., and Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *J. Exp. Child Psychol.* 111, 246–267. doi: 10.1016/j.jecp.2011.08.005

Bailey, D. H., Hoard, M. K., Nugent, L., and Geary, D. C. (2012). Competence with fractions predicts gains in mathematics achievement. *J. Exp. Child Psychol.* 113, 447–455. doi: 10.1016/j.jecp.2012.06.004

Barth, H. C., and Paladino, A. M. (2011). The development of numerical estimation: evidence against a representational shift. *Dev. Sci.* 14, 125–135. doi: 10.1111/j.1467-7687.2010.00962.x

Barth, H. C., Slusser, E., Cohen, D., and Paladino, A. M. (2011). A sense of proportion: commentary on Opfer, Siegler and Young. *Dev. Sci.* 14, 1205–1206. doi: 10.1111/j.1467-7687.2011.01081.x

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., and Zorzi, M. (2010). Numerical estimation in preschoolers. *Dev. Psychol.* 46, 545–551. doi: 10.1037/a0017887

Booth, J. K., and Newton, K. L. (2012). Fractions: could they really be the gatekeeper's doorman? *Contemp. Educ. Psychol.* 31, 247–253. doi: 10.1016/j.cedpsych.2012.07.001

Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 42, 189–201. doi: 10.1037/0012-1649.41.6.189

Booth, J. L., and Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Dev.* 79, 1016–1031. doi: 10.1111/j.1467-8624.2008.01173.x

Brannon, E. M., Wusthoff, C. J., Gallistel, C. R., and Gibbon, J. (2001). Numerical subtraction in the pigeon: evidence for a linear subjective number scale. *Psychol. Sci.* 12, 238–243. doi: 10.1111/1467-9280.00342

Burnham, K. P., Anderson, D. R., and Huyvaert, K. P. (2011). AIC model selection and multimodal inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* 65, 23–25. doi: 10.1007/s00265-010-1029-6

Cohen, D. J., and Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychon. Bull. Rev.* 18, 331–338. doi: 10.3758/s13423-011-0059-z

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N

Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol.* 122, 371–396. doi: 10.1037/0096-3445.122.3.371

Dehaene, S., and Cohen, L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex* 33, 219–250. doi: 10.1016/S0010-9452(08)70002-9

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239

De Hevia, M. D., Girelli, L., and Vallar, G. (2006). Numbers and space: a cognitive illusion? *Exp. Brain Res.* 168, 254–264. doi: 10.1007/s00221-005-0084-0

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., and Verschaffel, L. (2008). The relationship between the shape of the number line and familiarity with numbers in 5- to 9-year old children: evidence for a segmented linear model. *J. Exp. Child Psychol.* 99, 1–17. doi: 10.1016/j.jecp.2007.08.006

Ebersbach, M., Luwel, K., and Verschaffel, L. (2013). Comparing apples and pears in studies on magnitude estimations. *Front. Psychol.* 4:332. doi: 10.3389/fpsyg.2013.00332

Fischer, M. H. (2001). Number processing induces spatial performance biases. *Neurology* 57, 822–826. doi: 10.1212/WNL.57.5.822

Fischer, M. H. (2003). Spatial representations in number processing–evidence from a pointing task. *Vis. Cogn.* 10, 493–508. doi: 10.1080/13506280244000186

Fischer, U., Moeller, K., Bientzle, M., Cress, U., and Nuerk, H.-C. (2011). Sensori-motor spatial training of number magnitude representation. *Psychon. Bull. Rev.* 18, 177–183. doi: 10.3758/s13423-010-0031-3

Geary, D. C., Hoard, M. K., Nugent, L., and Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Dev. Neuropsychol.* 33, 277–299. doi: 10.1080/87565640801982361

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychol. Rev.* 84, 279–325. doi: 10.1037/0033-295X.84.3.279

Gibbon, J., and Church, R. M. (1981). Time left: linear versus logarithmic subjective time. *J. Exp. Psychol. Anim. Behav. Process.* 7, 87–108. doi: 10.1037/0097-7403.7.2.87

Gunderson, E. A., Ramirez, G., Beilock, S. L., and Levine, S. C. (2012). The relation between spatial skill and early number knowledge: the role of the linear number line. *Dev. Psychol.* 48, 1229–1241. doi: 10.1037/a0027433

Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). Language effects on children's non-verbal number line estimations. *J. Cross Cult. Psychol.* 42, 598–613. doi: 10.1177/0022022111406026

Hollands, J. G., and Dyre, B. P. (2000). Bias in proportion judgments: the cyclical power model. *Psychol. Rev.* 107, 500–524. doi: 10.1037/0033-295X.107.3.500

Hubbard, E. M., Piazza, M., Pinel, P., and Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nat. Rev. Neurosci.* 6, 435–448. doi: 10.1038/nrn1684

Hund, A. M., and Spencer, J. P. (2003). Developmental changes in the relative weighting of geometric and experience-dependent location cues. *J. Cogn. Dev.* 4, 3–38. doi: 10.1080/15248372.2003.9669681

Huttenlocher, J., Newcombe, N., and Sandberg, E. H. (1994). The coding of spatial location in young children. *Cogn. Psychol.* 27, 115–147. doi: 10.1006/cogp.1994.1014

Karolis, V., Iuculano, T., and Butterworth, B. (2011). Mapping numerical magnitudes along the right lines: Differentiating between scale and bias. *J. Exp. Psychol.* 140, 693–706. doi: 10.1037/a0024255

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., et al. (2011). Mental number line training in children with developmental dyscalculia. *Neuroimage* 57, 782–795. doi: 10.1016/j.neuroimage.2011.01.070

Laski, E. V., and Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Dev.* 78, 1723–1743. doi: 10.1111/j.1467-8624.2007.01087.x

LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Dev.* 81, 1753–1767. doi: 10.1111/j.1467-8624.2010.01508.x

LeFevre, J. A., Lira, C. J., Sowinski, C., Cankaya, O., Kamawar, D., and Skwarchuk, S. L. (2013). Charting the role of the number line in mathematical development. *Front. Psychol.* 4:641. doi: 10.3389/fpsyg.2013.00641

Moeller, K., and Nuerk, H.-C. (2011). Psychophysics of numerical representation: why seemingly logarithmic representations may rather be multi-linear. *Z. Psychol./J. Psychol.* 219, 64–70. doi: 10.1027/2151-2604/a000048

Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2009). Children's early mental number line: logarithmic or decomposed linear? *J. Exp. Child Psychol.* 103, 503–515. doi: 10.1016/j.jecp.2009.02.006

Moyer, R. S., and Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature* 215, 1519–1520. doi: 10.1038/2151519a0

Muldoon, K., Towse, J., Simms, V., Perra, O., and Menzies, V. (2013). A longitudinal analysis of estimation, counting skills, and mathematical ability across the first school year. *Dev. Psychol.* 49, 250–257. doi: 10.1037/a0028240

National Mathematics Advisory Panel. (2008a). *Foundations for Success: The Final Report of the National Mathematics Advisory Panel.* Washington, DC: US Department of Education.

National Mathematics Advisory Panel. (2008b). *Report of the Subcommittee on the National Survey of Algebra I Teachers.* Washington, DC: US Department of Education.

Petitto, A. L. (1990). Development of numberline and measurement concepts. *Cogn. Instr.* 7, 55–78. doi: 10.1207/s1532690xci0701_3

Ramani, G. B., and Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. *J. Appl. Dev. Psychol.* 32, 146–159. doi: 10.1016/j.appdev.2011.02.005

Restle, F. (1970). Speed of adding and comparing numbers. *J. Exp. Psychol.* 83, 274–278. doi: 10.1037/h0028573

Schneider, M., Heine, A., Thaler, V., Torbeyns, J., De Smedt, B., Verschaffel, L., et al. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cogn. Dev.* 23, 424–437. doi: 10.1016/j.cogdev.2008.07.002

Schutte, A. R., Simmering, V. R., and Ortmann, M. R. (2011). Keeping behavior in context: A dynamic systems account of a transition in spatial recall biases. *Spat. Cogn. Comput.* 11, 313–342. doi: 10.1080/13875868.2011.579212

Shaki, S., Fischer, M. H., and Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychon. Bull. Rev.* 16, 328–331. doi: 10.3758/PBR.16.2.328

Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi: 10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., et al. (2012). Early predictors of high school mathematics achievement. *Psychol. Sci.* 23, 691–697. doi: 10.1177/0956797612440101

Siegler, R. S., Fazio, L. K., Bailey, D. H., and Zhou, X. (2013). Fractions: the new frontier for theories of numerical development. *Trends Cogn. Sci.* 17, 13–19. doi: 10.1016/j.tics.2012.11.004

Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–243. doi: 10.1111/1467-9280.02438

Siegler, R. S., and Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschooler's numerical understanding. *J. Educ. Psychol.* 101, 545–560. doi: 10.1037/a0014239

Slusser, E. B., Santiago, R. T., and Barth, H. C. (2013). Developmental change in numerical estimation. *J. Exp. Psychol.* 142, 193–208. doi: 10.1037/a0028560

Spence, I. (1990). Visual psychophysics of simple graphical elements. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 683–692.

Sullivan, J. L., Juhasz, B. J., Slattery, T. J., and Barth, H. C. (2011). Adults' number-line estimation strategies: evidence from eye movements. *Psychon. Bull. Rev.* 18, 557–563. doi: 10.3758/s13423-011-0081-1

Whalen, J., Gallistel, C. R., and Gelman, R. (1999). Nonverbal counting in humans: the psychophysics of number representation. *Psychol. Sci.* 10, 130–137. doi: 10.1111/1467-9280.00120

Whyte, J. C., and Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschoolers. *Dev. Psychol.* 44, 588–596. doi: 10.1037/0012-1649.44.2.588

Zorzi, M., Priftis, K., and Umiltà, C. (2002). Neglect disrupts the mental number line. *Nature* 417, 138–139. doi: 10.1038/417138a

Two separate processes affect the development of the mental number line

Ronit Goldman[1]*, Joseph Tzelgov[1,2], Tamar Ben-Shalom[1] and Andrea Berger[1]

[1] Department of Psychology, Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[2] Department of Brain and Cognitive Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel
*Correspondence: ronit_goldman@hotmail.com

Two processes are apparently involved when adults compare magnitudes. One is an analog comparison process, which produces the **distance effect**—a decrease in reaction time (RT) the larger the difference between two compared magnitudes (Moyer and Landauer, 1967). The other is the activation of end stimuli (i.e., objects learned to be representing the smallest or the largest magnitudes in the set), which results in the **end effect**—faster processing of pairs that include the end stimuli of a set (Banks, 1977). Leth-Steensen and Marley (2000) proposed a formal model that shows how the two processes can account for comparisons RTs involving ordinal magnitudes.

Natural numbers are symbolic representations of magnitudes which, at least in adults, are apparently represented along a mental number line (e.g., Dehaene, 1997; Gallistel and Gelman, 2000). Automatic processing allows a direct retrieval of information stored in long-term memory (Logan, 1988; Perruchet and Vinter, 2002) and therefore can be used to examine the mental representation of numbers without contamination by intentionally applied strategies (Kallai and Tzelgov, 2009). This can be seen in Stroop-like phenomena when a task-irrelevant process affects processing of a relevant dimension (Tzelgov, 1997). Automatic processing of numbers can be accessed by a physical size comparison task, in which participants are presented with pairs of numbers differing in the numerical and physical size and instructed to select the physically larger number. The **size congruency effect** (SiCE), referring to faster RT for comparisons of pairs in congruent (e.g., 2_8) compared to incongruent (e.g., 2_8)

conditions (e.g., Henik and Tzelgov, 1982), serves as a marker of automaticity of numerical processing. Furthermore, linear increase of the SiCE as a function of intra-pair numerical distance is consistent with an analog representation of numbers (Tzelgov et al., 2013).

Pinhas and Tzelgov (2012) proposed that the two-process model of Leth-Steensen and Marley (2000) also applies to automatic processing of numbers. They attributed the monotonic increase of the SiCE with the intra-pair numerical distance (e.g., Henik and Tzelgov, 1982; Tzelgov et al., 2000) to the analog comparison process. In addition, the faster processing of pairs containing end stimuli was suggested to enlarge the SiCE due to earlier availability of numerical magnitude information (Schwarz and Ischebeck, 2003) and to attenuate the modulation of the effect by the numerical distance. This phenomenon was defined by Pinhas and Tzelgov (2012) as the automatic end effect (AEE), and was assumed to result from real-world experience. The effect shown to exist for 0, and for 1 in the absence of 0, but not for larger numbers; that is, the effect was absent when 2 was the smallest number in the set. This finding is important as it shows the special status of 1 (and 0) as the semantically smallest number stored in long-term memory (Tzelgov et al., 2013) and is consistent with the special status of 1 as hypothesized by Leslie et al. (2008).

While the picture is relatively clear with regard to number representation in adults, less is known about such representation in children. If its emergence reflects learning (e.g., Verguts and Fias, 2008), both processes may be involved in the formation of the mental number line. In particular, we

do not know how the processes involved in number comparison (analog number comparison and mapping 1 as the smallest number) develop in children and contribute to the emergence of the mental number line. Thus, the development of these processes is of interest.

Several studies have investigated numerical comparisons in children and used the SiCE to learn about the development of automatization of numerical processing. Rubinsten et al. (2002) reported a numerical distance effect in numerical judgments in kindergarteners (see also Sekuler and Mierkiewicz, 1977) but the SiCE in physical comparisons emerged by the end of first grade, with no modulation by numerical distance. Girelli et al. (2000) classified pairs of numbers as "unilateral" (both numbers smaller or both larger than five) and "bilateral" (one number smaller and the other larger than 5). In numerical comparisons, laterality, being positively correlated with distance, affected latencies in first, third, and fifth graders. In this study, the SiCE was found for third and fifth graders but not for first graders. Zhou et al. (2007) were the only ones to have shown an SiCE modulated by intra-pair numerical distance for Chinese kindergartners, consistent with the assumption of the arrangement of numbers along the mental number line. They attributed the emergence of the effect at this relatively early age to cultural differences.

In the current work we examined the effects of the two comparison processes in both intentional and automatic numerical judgments of children. We used the data of 118 kindergarteners (Mean = 6.1 years, SD = 4.2 months) from the study of

Ben-Shalom et al. (submitted), where the authors presented children with numerical and physical comparison tasks on stimuli differing in numerical values and physical sizes. Number pairs were presented with intra-pair numerical distance of 1 (the pairs: 1_2, 3_4, 6_7, 8_9), 2 (the pairs: 1_3, 2_4, 6_8, 7_9), and 5 (the pairs: 1_6, 2_7, 3_8, 4_9). Each number in a pair appeared an equal number of times on each side of the screen center. In both tasks, 24 congruent pairs (e.g., 3_8), and 24 incongruent pairs (e.g., 3_8) were created by presenting numbers in physical sizes of 10 mm (small size) or 13 mm (large size). To demonstrate the two processes in both tasks, we conducted separate analyses for pairs that contain 1, and for pairs that did not contain 1. In all analyses the nominal significance level was defined as $p < 0.05$.
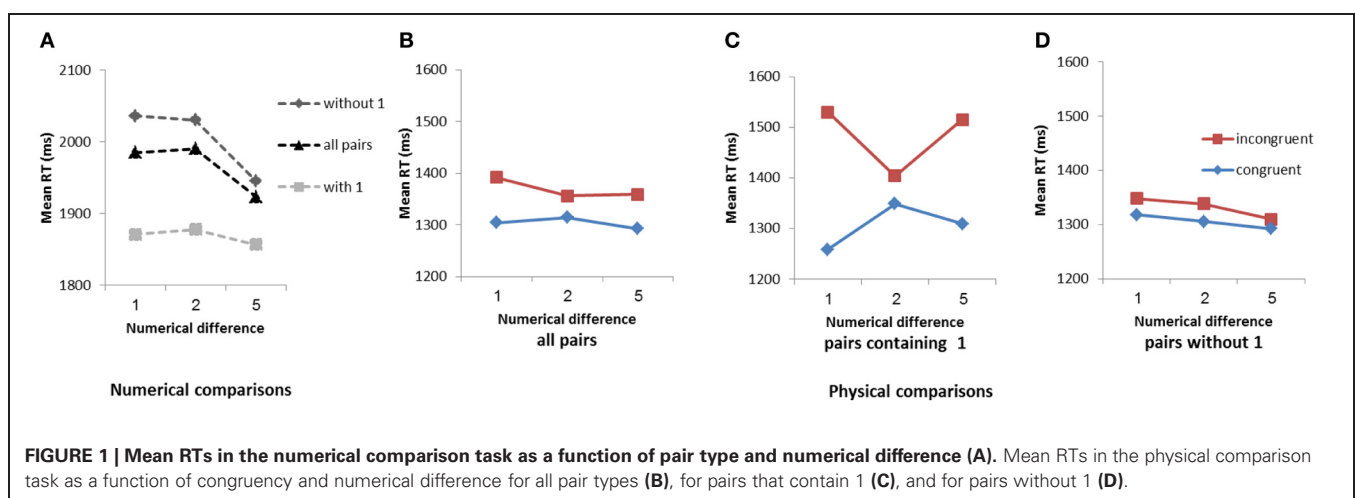
In the numerical comparison task (**Figure 1A**), when pairs that contained 1 and pairs without 1 were included in a common analysis, the children showed a clear distance effect [$F_{(1, 117)} = 11.29$, $MSE = 26,282$, $\eta_p^2 = 0.09$] (see Ben-Shalom et al., submitted). However, pairs containing 1 were compared faster than comparisons of pairs that did not contain 1 [$F_{(1, 117)} = 46.67$, $MSE = 70,729$, $\eta_p^2 = 0.28$]. Furthermore, comparisons of pairs including 1 showed no sensitivity to numerical distance ($F < 1$). In contrast, comparisons of pairs without 1 were faster with the increase in the numerical distance between the numbers in the pair, as indicated by the linear trend for numerical difference (1, 2, and 5) [$F_{(1, 117)} = 17.05$, $MSE = 34,489, \eta_p^2 = 0.13$].

In the physical comparison task, the SiCE computed for pairs with and without 1 (**Figure 1B**) was found to be significant [$F_{(1, 117)} = 28.52$, $MSE = 26,253$, $\eta_p^2 = 0.20$] and was not modulated by numerical distance ($F < 1$). In line with the notion of the AAE, the size congruity effect in the physical comparison task was much larger in pairs that contained 1 than in pairs without 1 [$F_{(1, 115)} = 19.23$, $MSE = 104,335$, $\eta_p^2 = 0.14$] (compare **C** and **D** in **Figure 1**). In fact, the SiCE was apparent only for pairs containing 1 [$F_{(1, 115)} = 34.06$, $MSE = 166,254$, $\eta_p^2 = 0.23$], with no evidence of linear modulation (**Figure 1C**), and was minimal and marginally significant for pairs without 1 [$F_{(1, 117)} = 3.85$, $MSE = 32,329$, $\eta_p^2 = 0.03$] (**Figure 1D**). The SiCE for pairs containing 1 did not differ for distances 1 and 5 ($F < 1$) and was larger for the distances of 1 and 5 than for the distance of 2 [$F_{(1, 115)} = 10.77$, $MSE = 111,129, \eta_p^2 = 0.09$]. Importantly, an analysis performed on the largest number in the set found no indication for the automatic processing of 9 as an end stimulus, the SiCE for comparisons of pairs containing 9 was non-significant ($F < 1$).

The present study demonstrates the distance effect and the end effect in kindergarteners, showing that in young children as in adults, an analog comparison process and mapping to end anchors are involved in magnitude comparisons (Leth-Steensen and Marley, 2000). Our results replicate the distance effect in kindergarteners (e.g., Sekuler and Mierkiewicz, 1977; Rubinsten et al., 2002) but only

in comparisons that did not contain 1. Consistent with Zhou et al. (2007), we also found an SiCE in kindergarteners. The effect was enlarged for pairs including 1, as found for adults (Pinhas and Tzelgov, 2012), and did not increase with numerical distance. In pairs that did not include 1, the SiCE was minimal and insensitive to the numerical distance. It follows that the AEE for 1 and the modulation of the SiCE by intra-pair numerical distance become automatized during development at different rates. Because automaticity is achieved with experience, the fact that only the processing of 1 as an end stimulus affected automatic numerical comparisons suggests that this process develops earlier than the analog comparison process. The finding that 1, but not 9, showed an AEE further implies that 1 has a special status as the smallest member of the mental number line. As kindergartners acquire real-world experience with numbers larger than 9 (e.g., 10) the absence of the effect for 9 may result from such experience. In that sense, it is similar to the absence of the AEE for 2 when it was the smallest number in the set, as reported by Pinhas and Tzelgov (2012).

Leslie et al. (2008) refer to the special role of 1 in the generation of natural numbers. They proposed that humans are born with (1) the ability of symbolic representation of (at least) the minimal possible magnitude by a numeral equivalent to 1, and (2) the function "next," which recursively allows adding 1 to each (natural) number, thus enabling to generate the representation of each and every natural number.

FIGURE 1 | Mean RTs in the numerical comparison task as a function of pair type and numerical difference (A). Mean RTs in the physical comparison task as a function of congruency and numerical difference for all pair types **(B)**, for pairs that contain 1 **(C)**, and for pairs without 1 **(D)**.

The special status of the number 1 results from the recursive rule, as it is the only number that can be used to generate a representation for each and every natural number in this manner. Our results showing that kindergartners, like adults, show an AEE for 1 (Pinhas and Tzelgov, 2012), implying they process 1 as the smallest magnitude, is in line with Leslie et al.'s assumptions. Assuming the mental number line is created by a next function of the smallest magnitude 1, the formation of the order relations between magnitudes is supposedly created by learning of adjacent pairs. The first order relation to be learned is that the magnitude 1 is smaller than the magnitude 2. In accordance with this suggestion is our finding that in comparisons of pairs containing 1, an enlarged SiCE was found for the numerical distance of 1, that is, comparisons of the pair 1_2.

The special processing of 1 as the smallest magnitude can account for some of the differences found in studies of numerical processing development. Because comparisons of pairs containing 1 show an increased SiCE, the effect may be manipulated by inclusion or exclusion of 1 in the analysis, as demonstrated in this study. In line with our suggestion, Zhou et al. (2007), who found an SiCE for kindergartners, included the number 1 in their stimuli set (in a third of the experiment trials), whereas studies that did not include 1 in the stimuli set showed the emergence of the SiCE only later at the end of first grade (Rubinsten et al., 2002) or in third grade (Girelli et al., 2000).

The insensitivity of the SiCE to numerical distance implies kindergartners do not have an analog representation of numerals as magnitudes in the long-term memory. This does not necessarily mean that magnitudes are not represented mentally along a mental line, but rather the mapping of symbols to this representation is not automatic at this early age.

The earlier development of mapping to end anchors as compared with the analog comparison process is also evident in studies in which the association between symbols and magnitudes is artificially created and learned (e.g., Riley and Trabasso, 1974; Banks, 1977; Tzelgov et al., 2000). The linear ordering of a set is constructed from the ends inward, as participants first learn to identify the end stimuli of the set, and gradually fill in the order relations of stimuli from the rest of the set (e.g., Riley and Trabasso, 1974).

SUMMARY

The current study showed intentional and automatic numerical judgments of kindergartners are affected by an analog comparison process and the processing of end stimuli. The number 1 was found to have a special status as the smallest number, as implied by both intentional and AEEs. The distance effect was found in intentional comparisons of numbers but was absent in automatic processing. These results indicate the processing of end stimuli develop earlier than the analog comparison process. Finally, we demonstrated the inclusion of 1 in the stimuli set increases the SiCE and suggested this can account for the emergence or absence of the effect as reported in the literature.

ACKNOWLEDGMENTS

REFERENCES

Banks, W. P. (1977). "Encoding and processing of symbolic information in comparative judgments," in *The Psychology of Learning and Motivation* ed G. Bower (New York, NY: Academic Press), 101–159.

Dehaene, S. (1997). *The Number Sense*. New York, NY: Oxford University Press.

Gallistel, C. R., and Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends Cogn. Sci.* 4, 59–65. doi: 10.1016/S1364-6613(99)01424

Girelli, L., Lucangeli, D., and Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *J. Exp. Child Psychol.* 76, 104–122. doi: 10.1006/jecp.2000.2564

Henik, A., and Tzelgov, J. (1982). Is three greater than five: the relation between physical and semantic size in comparison tasks. *Mem. Cognit.* 10, 389–395. doi: 10.3758/BF03202431

Kallai, A. Y., and Tzelgov, J. (2009). A generalized fraction: an entity smaller than one on the mental number line. *J. Exp. Psychol. Hum.* 35, 1845–1864. doi: 10.1037/a0016892

Leslie, A. M., Gelman, R., and Gallistel, C. (2008). The generative basis of natural number concepts. *Trends Cogn. Sci.* 12, 213–218. doi: 10.1016/j.tics.2008.03.004

Leth-Steensen, C., and Marley, A. A. J. (2000). A model of response time effects in symbolic comparison. *Psychol. Rev.* 107, 62–100.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychol. Rev.* 95, 492–527. doi: 10.1037/0033-295X.95.4.492

Moyer, R. S., and Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature* 215, 1519–1520. doi: 10.1038/2151519a0

Perruchet, P., and Vinter, A. (2002). The self-organizing consciousness. *Behav. Brain Sci.* 25, 297–329.

Pinhas, M., and Tzelgov, J. (2012). Expanding on the mental number line: zero is perceived as the "smallest". *J. Exp. Psychol. Learn.* 38, 1187–1205. doi: 10.1037/a0027390

Riley, C. A., and Trabasso, T. (1974). Comparatives, logical structures, and encoding in a transitive inference task. *J. Exp. Child Psychol.* 17, 187–203. doi: 10.1016/0022-0965(74)90065-4

Rubinsten, O., Henik, A., Berger, A., and Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *J. Exp. Child Psychol.* 81, 74–92. doi: 10.1006/jecp.2001.2645

Schwarz, W., and Ischebeck, A. (2003). On the relative speed account of number-size interference in comparative judgments of numerals. *J. Exp. Psychol. Hum. Percept. Perform.* 29, 507–522. doi: 10.1037/0096-1523.29.3.507

Sekuler, R., and Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Dev.* 48, 630–633. doi: 10.2307/1128664

Tzelgov, J. (1997). Specifying the relations between automaticity and conscious: a theoretical note. *Conscious Cogn.* 6, 441–451. doi: 10.1006/ccog.1997.0303

Tzelgov, J., Ganor-Stern, D., Kallai, A. Y., and Pinhas, M. (2013). "Primitives and non-primitives of numerical representations," in *Oxford Handbook of Numerical Cognition* eds R. Cohen Kadosh, and A. Dowker (Oxford: Oxford Academic Press), 55–66.

Tzelgov, J., Yehene, V., Kotler, L., and Alon, A. (2000). Automatic comparisons of artificial digits never compared: learning linear ordering relations. *J. Exp. Psychol. Learn.* 26, 103–120. doi: 10.1037/0278-7393.26.1.103

Verguts, T., and Fias, W. (2008). Symbolic and nonsymbolic pathways of number processing. *Philos. Psychol.* 21, 539–554. doi: 10.1080/09515080802285545

Zhou, X., Chen, Y., Chen, C., Jiang, T., Zhang, H., and Dong, Q. (2007). Chinese kindergartners' automatic processing of numerical magnitude in Stroop-like tasks. *Mem. Cognit.* 35, 464–470. doi: 10.3758/BF03193286

How number-space relationships are assessed before formal schooling: A taxonomy proposal

Katarzyna Patro[1,2]*, Hans-Christoph Nuerk[2,3], Ulrike Cress[2] and Maciej Haman[1]

[1] Department of Psychology, University of Warsaw, Warsaw, Poland
[2] Knowledge Construction Lab, Knowledge Media Research Center, Tübingen, Germany
[3] Department of Psychology, University of Tübingen, Tübingen, Germany

The last years of research on numerical development have provided evidence that spatial-numerical associations (SNA) can be formed independent of formal school training. However, most of these studies used various experimental paradigms that referred to slightly different aspects of number and space processing. This poses a question of whether all SNAs described in the developmental literature can be interpreted as a unitary construct, or whether they are rather examples of different, but related phenomena. Our review aims to provide a starting point for a systematic classification of SNA measures used from infancy to late preschool years, and their underlying representations. We propose to distinguish among four basic SNA categories: (i) cross-dimensional magnitude processing, (ii) associations between spatial and numerical intervals, (iii) associations between cardinalities and spatial directions, (iv) associations between ordinalities and spatial directions. Such systematization allows for identifying similarities and differences between processes and representations that underlie the described measures, and also for assessing the adequacy of using different SNA tasks at different developmental stages.

Keywords: infants, number, preschoolers, space, SNA, SNARC, taxonomy

INTRODUCTION

Over the last 20 years, different experimental paradigms have been employed to examine mental connections between numbers and space. For instance, a parity judgment task with bimanual responses usually reveals the SNARC effect (Spatial-Numerical Associations of Response Codes) which is defined as left-sided response advantage for smaller numbers, and right-sided for larger numbers (Dehaene et al., 1993). This and other kinds of spatial-numerical associations (SNA) were originally thought to be acquired through a long-term experience in a certain cultural environment with specific script direction and formal school training. However, this late development account has been recently challenged because some forms of number-space associations have been discovered to develop before formal schooling.

A closer look reveals that the various SNA investigations in young children have referred to different associations between numbers and space. In fact, the term "space" in number-space relationship is understood and assessed in a variety of ways. Sometimes its directionality is examined (left-right dimension), whereas in other studies its non-directional extension (number lengths). The term "number" has also been used with different connotations: Some studies referred to ordinality (counting) and others to cardinality (numerical magnitudes).

In sum, several theoretical and methodological differences between SNA tasks exist in studies of children, which may address different underlying space and number representations. To date, all SNA studies have been lumped into one category, but no systematic distinction of early SNA measures exists. In the current review, a systematic taxonomy is proposed to enable classifying present and future approaches to research on SNA before school education.

DISTINGUISHING DIFFERENT SNA EFFECTS IN PRESCHOOLERS AND INFANTS: TOWARD A TAXONOMY

While many comprehensive reviews have described varieties of SNAs and their assessments in adults (Gevers and Lammertyn, 2005; Hubbard et al., 2005; Wood et al., 2008), there is only one theoretical article on the pre-linguistic basis of the number-space link (de Hevia et al., 2012), which includes a short overview about developmental studies. While this review is ground-breaking in that it establishes early SNA as a widespread phenomenon, it does not provide a systematization of employed tasks, nor the underlying representations assessed in different age ranges. In the following paragraphs, we emphasize the reasons why a systematization of SNA tasks is important, especially for developmental studies.

First, it is a classical question in psychology, whether different tasks used to assess an underlying construct measure the same construct. The broader the construct is, the more critical this question becomes. The SNA is an example of such a broad construct. Different number representations and their components can be related to space (non-symbolic numerosities, symbolic magnitude, position in a sequence), and different aspects of space can be studied in relation to number processing (length, directionality). In adult studies, not all of these components comprising number-space associations activate the same neurocognitive representations (Turconi et al., 2004; Cohen

Kadosh et al., 2007). Therefore, these components should be systematically distinguished.

Second, adult reviews are not sufficient, because different SNAs might emerge at different developmental stages, as different aspects of numerical or spatial information are accessible at different ages. Therefore, general conclusions from adult SNA paradigms may fall short, because the numerical or spatial information employed is inaccessible to young children.

For these reasons, in this review we outline systematic distinctions between SNA effects and their assessment in children before school age. We show that the methods used so far might be based on different aspects of number and space processing—on directional and non-directional space representation, and the representation of cardinal, ordinal or interval numerical information. Because this variety of representations might lead to different ways of defining a number-space link, we suggest distinguishing among at least four main SNA categories, as a first step. Two of them refer to a non-directional number-space mapping: (i) cross-dimensional magnitude processing, and (ii) associations between spatial and numerical intervals. The other two categories refer to directional number-space mapping: (iii) associations between cardinalities and spatial directions, and (iv) associations between ordinalities and spatial directions.

The order of these categories does not reflect the order in which they arise across lifespan. As will be shown later, investigations within some of these categories are focused mostly on one age group and are based on similar experimental paradigms, developed by one or two research groups. Therefore, the current state of research does not allow making strong conclusions about the origins and developmental trajectory of SNA.

Furthermore, we are aware that other distinctions beyond these four categories are possible. For instance, one might distinguish between symbolic and non-symbolic numerosities. However, although symbolic numbers have been employed in some studies reported here, using them as stimuli might be sometimes misleading because only the oldest children master them reliably (see the discussion of the last category). Therefore, we have constrained our categories to those distinctions that have played a major role in prior research on SNA in young children. Thus, our four categories provide a starting point based on the research, but are not meant to preclude additional distinctions that might become relevant in future studies.

Finally, for most of these tasks, it is discussed whether the observed number-space associations are caused by respective representations or rather by corresponding task properties. Although it is beyond a scope of this review to discuss thoroughly all the critics, we will refer to them shortly to provide the reader with a critical overview of the SNA research.

In the next sections, we provide a short outline of assessments and representations tested in each SNA category, the age of the participants with which they have been applied, and describe specific features of the representations and their underlying processes. We also briefly refer to certain controversies aroused by some of these measures. This systematization is summarized in **Table 1**.

SNA CATEGORY 1: CROSS-DIMENSIONAL MAGNITUDE PROCESSING

MEASURES

Extraction of an abstract rule across spatial-numerical dimensions (de Hevia and Spelke, 2010; Lourenco and Longo, 2010). Eight- and nine-month-old infants were habituated to an abstract rule referring either to numerosity or to length (e.g., color pattern cues assigned to a certain magnitude or ascending/descending orders of magnitude sequences). In a testing phase, a magnitude dimension presented during habituation (e.g., numerosity) was replaced by another dimension (e.g., length). Despite this change, infants detected if an abstract rule, to which they were habituated, had been reversed.

Numerosity-length matching (de Hevia and Spelke, 2010; de Hevia et al., 2012). Four-year-old preschoolers were presented with a rule of a positive magnitude matching (larger numerosities paired with longer lines), or an inverse rule (larger-shorter pairings). Most of the children found a correct match in the testing trials, but only in the positive mapping condition. Similarly, following a short familiarization to larger-numerosity-longer-line pairings, 8-month-old infants demonstrated a looking time preference for novel stimuli matched according to the same, recently acquired rule. A familiarization to an inverse rule did not cause any preference in later trials.

Line bisection (de Hevia and Spelke, 2009). Five-year-old children were instructed to mark the center of a horizontal line flanked by two arrays of dots with different numerosities. Children placed the midpoint of a line closer to a larger numerosity side.

MAIN FEATURES OF THE SNA CATEGORY 1

The above tasks used different experimental paradigms to explore interrelations between cardinal aspects of non-symbolic numerosities and non-directional spatial dimension - line length. They indicate that length and numerosity are related even for infants and young children, either in a complementary way, carrying together a common abstract rule, or in an interfering way, when a value of one dimension modulates processing of another dimension.

Such kinds of interactions can be explained by theories postulating a generalized system for processing magnitudes of different sorts (Walsh, 2003; Cantlon et al., 2009). Processing of length, number, and other magnitude dimensions share many representational features, for example a continuous and quantitative metric or early ontogenetic and evolutionary beginnings. Therefore, it is postulated that the human mind is predisposed to form a common representational framework for numerical and spatial quantities, or at least to map one dimension onto another. As we will see, this makes the numerosity-length bond unique and qualitatively different from the other SNA categories, for which an influence of cultural factors has been postulated or demonstrated.

These methods are also critically discussed. When stimuli consist of object's collections, certain perceptual set features (density, elements' size) correlated with numerosities may drive participants' choices. For instance, Gebuis and Gevers (2011) challenged the numerical explanation of the line bisection bias. In their experiment, when numerosity was negatively correlated with a

Table 1 | The four distinguished SNA categories, their measures, studies in which these measures were used, and mean age of participants.

Category of spatial-numerical associations (SNA)	Task[a]	Study	Mean age (average ages provided as decimal fractions have been recalculated to a "years-months-days" format)
SNA1: Cross-dimensional magnitude processing	Extraction of an abstract rule across spatial-numerical dimensions	Lourenco and Longo, 2010 de Hevia and Spelke, 2010, Exp. 1	$M = 0$ year 9 months 4 days $M = 0$ year 8 months
Space: non-directional Number: cardinal	Numerosity-length matching	de Hevia and Spelke, 2010, Exp. 2, 3 de Hevia et al., 2012;	$M = 0$ year 8 months 3 days (Exp. 2), $M = 0$ year 8 months 2 days (Exp. 3) $M = 4$ years 1 month
	Line bisection	de Hevia and Spelke, 2009, Exp. 4, 5, 6	$M = 5$ years 1 months (Exp. 4), $M = 4$ years 11 months (Exp. 5), $M = 4$ years 9 months (Exp. 6)
SNA2: Associations between spatial and numerical intervals	Number line task	Siegler and Booth, 2004, Exp. 1, 2	$M = 5$ years 10 months (Exp. 1), $M = 6$ years 1 month (Exp. 2)
		Booth and Siegler, 2006, Exp. 1 Ebersbach et al., 2008 Berteletti et al., 2010, Exp. 1, 2	$M = 5$ years 8 months $M = 5$ years 3 months $M_1 = 4$ years 0 month, $M_2 = 5$ years 0 month, $M_3 = 5$ years 11 months (Exp. 1) $M_1 = 4$ years 0 month, $M_2 = 4$ years 11 months, $M_3 = 5$ years 10 months (Exp. 2)
Space: non-directional[b] Number: interval		Barth and Paladino, 2011, Exp. 2 Fischer et al., 2011 Muldoon et al., 2011 Slusser et al., 2013, Exp. 1 Ebersbach, in press	$M = 5$ years 7 months $M = 5$ years 10 months $M_1 = 4$ years 6 months, $M_2 = 5$ years 2 months $M = 5$ years 11 months $M = 6$ years 0 month
SNA3: Associations between cardinalities and spatial directions	Numerosity comparison	Patro and Haman, 2012	$M = 4$ years 0 month
Space: directional Number: cardinal	SNARC tasks	Hoffmann et al., 2013 Ebersbach et al., 2013	$M_1 = 5$ years 6 months, $M_2 = 5$ years 10 months $M = 5$ years 11 months
SNA4: Associations between ordinalities and spatial directions	Counting	Opfer and Thompson, 2006 (Study1) Opfer and Furlong, 2011 Shaki et al., 2012, Exp. 2	$M_1 = 3$ years 3 months, $M_2 = 4$ years 0 month, $M_3 = 5$ years 2 months $M_1 = 4$ years 6 months, $M_2 = 4$ years 6 months, $M_3 = 4$ years 6 months, $M_4 = 4$ years 5 months $M_1 = 4$ years 2 months, $M_2 = 4$years 1 months, $M_3 = 3$ years 11 months
Space: directional Number: ordinal	Addition/removal of one object	Opfer and Thompson, 2006 (Study1) Opfer and Furlong, 2011	$M_1 = 3$ years 3 months, $M_2 = 4$ years 0, $M_3 = 5$ years 2 months $M_1 = 4$ years 6 months, $M_2 = 4$ years 5 months, $M3 = 4$ years 6 months, $M_4 = 4$ years 5 months
	Spatial search	Opfer et al., 2010, Exp. 1 Opfer and Furlong, 2011	$M_1 = 4$ years 6 months, $M_2 = 4$ years 6 months $M_1 = 4$ years 6 months, $M2 = 4$ years 6 months, $M_3 = 4$ years 6 months, $M_4 = 4$ years 5 months

[a]Both ordinal and interval numerical information were studied in relation to space also with a graphic production task (Tversky et al., 1991). Five- and 6-year-old children placed on a paper two stickers representing larger or smaller amounts of objects. These stickers had to be placed in relation to the third, centrally positioned sticker that represented a medium amount of the same things. However, interval relations between numerosities were not mapped by children as spatial distances, and there were no consistent directional biases.

[b]One recent study (Ebersbach, in press) introduced additional manipulation to the original number-line task by changing the direction of a displayed line (left-to-right vs. right-to-left). Processing of numerical and spatial (non-directional) intervals was still a main representation activated, however, it was affected by another aspect of space processing—its directionality. It shows that our spatial categories are at least in this one case overlapping and future studies might require extention of our taxonomy.

set's area, a bias occurred toward a smaller numerosity (larger area). They argued that the bisection bias was visually rather than numerically driven. However, several arguments against their critics were also presented (de Hevia, 2011). Whether children process numerosities independently of perceptual cues becomes now an intensively discussed issue (Cantrell and Smith, 2013; Szűcs et al., 2013; Libertus et al., 2014 for opposite views), which will certainly trigger more studies addressing this question.

SNA CATEGORY 2: ASSOCIATIONS BETWEEN SPATIAL AND NUMERICAL INTERVALS
MEASURE
Number-line task (Siegler and Booth, 2004; Booth and Siegler, 2006; Ebersbach et al., 2008; Berteletti et al., 2010; Barth and Paladino, 2011; Fischer et al., 2011; Muldoon et al., 2011; Ebersbach, in press). Preschoolers (from the age of 4 and older) were instructed to place a target number on a line with 0 or 1 at one end, and 10, 20, 100 or other numbers at the second end (depicted in Arabic or non-symbolic format). Lower accuracy in younger children has been observed, characterized by overestimation of interval distances between smaller numbers (logarithmic scaling). Accuracy in this task increases with age and is correlated with other numerical capabilities.

MAIN FEATURES OF THE SNA CATEGORY 2
This task requires representation of interval scaling instead of cardinal magnitudes. If it is solved correctly, the interval of the numerical magnitude is isomorphic to spatial magnitude of the presented line. Thus, space representation activated here is non-directional (like in SNA1).

A second important feature is that this task enforces spatial mapping of numbers. Thus, its primary purpose should be distinguished from the purposes of SNA1 measures: It does not evaluate whether numbers can be spontaneously related to space, but rather whether such a relationship, if it is induced, is constructed in a systematic interval-scaled way.

The idea that the estimations of the number-line task reflect scaling of a mental number line representation in an isomorphic way (Siegler and Opfer, 2003) has been criticized by various authors. Some authors argue that a logarithmic fit is produced by multi-linear representations for different number ranges (Ebersbach et al., 2008; Moeller et al., 2009). Other suggest that different proportional strategies are applied to solve the task (Barth and Paladino, 2011). The mental processes and strategies underlying the number-space associations in this task are thus still controversial.

SNA CATEGORY 3: ASSOCIATIONS BETWEEN CARDINALITIES AND SPATIAL DIRECTIONS
MEASURE
Numerosity comparison (Patro and Haman, 2012). Four-years-old children compared numerosities of two bilaterally presented sets. Reactions were faster when a smaller set was on the left side, and a larger one on the right side.

SNARC tasks (Ebersbach et al., 2013; Hoffmann et al., 2013). Five-years-old children classified a single number (or set of dots) between two categories (red/green, smaller/larger than x) using left and right buttons. Left-sided reactions were faster to smaller numbers, and right-sided to larger numbers.

MAIN FEATURES OF THE SNA CATEGORY 3
SNA3 and SNA4 differ in a major way from SNA1 and 2, because these associations are about spatial directionality and not about non-directional spatial magnitude. Larger numbers are generally associated with one side of horizontal space (the right side in Western culture), while smaller numbers are associated with the other. In the above tasks, such a directional representation is formed for cardinal numbers.

There are at least two critical issues for this category. First, it is controversial whether an adult-like association is built between numerical and spatial representations (Dehaene et al., 1993), or rather between "smaller/larger" and "left/right" verbal categories (Nuerk et al., 2004; Proctor and Cho, 2006; Gevers et al., 2010). Which kind of association preschoolers build, has not yet been examined. Second, reading direction was traditionally believed to be a main source of SNARC (Zebian, 2005; Shaki and Fischer, 2008), whereas other not determined pre-literate factors may build SNA in preschoolers. If such factors are among early acquired cultural skills, it would strengthen the difference between SNA1 and SNA3 because the first category may be more innately determined.

SNA CATEGORY 4: ASSOCIATIONS BETWEEN ORDINALITIES AND SPATIAL DIRECTIONS
MEASURES
Counting (Opfer and Thompson, 2006; Opfer and Furlong, 2011; Shaki et al., 2012). English-, Hebrew-, and Arabic-speaking preschoolers (age 3–6) were asked to count objects aligned horizontally in a row. Most of English-speaking children counted from left to right, whereas Hebrew- and Arabic-speaking children counted from right to left.

Addition/removal of one object (Opfer and Thompson, 2006; Opfer and Furlong, 2011). In an addition task, English-speaking preschoolers (age 3–5) were presented with three objects aligned in a row, and asked to add one object. In a subtraction task, they were presented with four objects and asked to remove one of them. The number of participants who added to and removed from the correct right end increased with age, but it was still below 50%.

Spatial search (Opfer et al., 2010; Opfer and Furlong, 2011). Four-year-old English-speaking children were presented with two boxes containing seven compartments, numbered in a left-to-right or right-to-left order. An experimenter showed a hidden star in one compartment of the first box, and asked children to find a similar star in the second box, in the compartment with the same number. Children performed more accurately in a condition when compartments were numbered from left to right than in a condition with the reversed numerical order.

MAIN FEATURES OF THE SNA CATEGORY 4
The tasks from this category are based on directional processing of numerical orders. Ordering tasks may be subdivided into two categories: One category without necessary access to magnitude (e.g., counting tasks) and another category, in which

order necessarily involves magnitude access or cardinality. For instance, ascending/descending numerosity-length matching (de Hevia and Spelke, 2010) requires some sort of order representation which is based on magnitude (smaller comes before larger). Therefore, in this category, we suggest to include only those studies in which a position of a number in a sequence (four comes before five, rightmost element should be removed), and not its magnitude, is relevant for solving a task.

Research with adults suggests that ordinal and cardinal processing of numbers rely at least partly on different processes. For instance, neural activation observed during classification of numbers as smaller or larger than a target has a different spatio-temporal pattern than during classifying them as coming before or after a target (Turconi et al., 2004). There is a strong agreement that both kinds of numerical processes constitute a number semantics (Sury and Rubinsten, 2012), but their contributions to building a numerical representation, especially in children, might be different.

In young preschoolers, the fixed order of a counting list is not primarily a numerical representation. Access to numerical semantics through symbolic notation establishes itself through the whole preschool period (Wynn, 1990; Le Corre and Carey, 2007). Although 3-year-olds might recite sequences of number words, they are usually not yet able to assign all learned numerals to corresponding values or to use a counting procedure to establish a set numerosity.

These considerations make the conceptual distinction between cardinal tasks (SNA3) and ordinal tasks (SNA4) valid, but a distinction does not imply that ordinal and cardinal knowledge are unrelated—the fixed order of a counting list may provide a scaffold for later understanding of cardinality. According to the cardinal principle of counting, the magnitude of the last ordinal element equals the cardinality. Children realize that quite late, usually only around the age of four. Once children have realized the relation between symbolic number order and number cardinality, however, order tasks like a counting or digit search task may become theoretically ambiguous, because it is sometimes not entirely clear to what extent ordinal and cardinal aspects of numbers contribute to the observed spatial associations.

Another important feature of SNA4 refers to its validity. In the tasks from this category, object stimuli are already arranged spatially forming a row. Thus, alike in SNA2, and contrary to SNA1 and 3, this task induces spatial coding and cannot evidence spontaneous number-space mapping. Note, however, that only a spatial location of response is enforced, but not directionality, because children can also count or subtract from the middle.

SUMMARY

In the taxonomy proposed here, we have distinguished four basic categories of SNA research in children before school age. Some of them share basic features, like directional components, or using a spatial magnitude as a basic mapping metric, but they differ at the same time in other aspects (processing ordinal, cardinal or interval number magnitudes).

These four distinguished SNA categories, although based on different number and space representations, are not intended to be exclusive. Our aim was to accentuate that although they are

overlapping and related to each other, their basic mechanisms, developmental trajectory, and susceptibility to cultural experience might be different.

These categories are a starting point for categorizing the most important distinctions in current preschool SNA research. They may not be comprehensive for all future results. For instance, we have outlined that the distinction between symbolic and non-symbolic magnitudes is not of major importance in prior developmental studies on SNA, but it is possible that future studies will bring new insights about the symbol-grounding problem that will require extension of this taxonomy.

Developmental studies on SNA are still sparse and other research paradigms might bring new insights into individual trajectories, new distinctions, and the causal relations of different number-space associations. They could also help to clarify the issues of validity of different SNA measures. However, we believe that the current taxonomy helps to understand that the growing evidence about SNA before schooling is not unitary, and should be categorized with consideration of the major distinctions of underlying space and number concepts used in empirical studies to date.

REFERENCES

Barth, H. C., and Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Dev. Sci.* 14, 125–135. doi: 10.1111/j.1467-7687.2010.00962.x

Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., and Zorzi, M. (2010). Numerical estimation in preschoolers. *Dev. Psychol.* 46, 545–551. doi: 10.1037/a0017887

Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 42, 189–201. doi: 10.1037/0012-1649.41.6.189

Cantlon, J. F., Platt, M. L., and Brannon, E. M. (2009). Beyond the number domain. *Trends Cogn. Sci.* 13, 83. doi: 10.1016/j.tics.2008.11.007

Cantrell, L., and Smith, L. B. (2013). Open questions and a proposal: a critical review of the evidence on infant numerical abilities. *Cognition* 128, 331–352. doi: 10.1016/j.cognition.2013.04.008

Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* 122, 371–396.

de Hevia, M.-D. (2011). Sensitivity to number: reply to Gebuis and Gevers. *Cognition* 121, 253–255. doi: 10.1016/j.cognition.2011.04.002

de Hevia, M.-D., Girelli, L., and Cassia, V. M. (2012). Minds without language represent number through space: origins of the mental number line. *Front. Psychol.* 3:466. doi: 10.3389/fpsyg.2012.00466

de Hevia, M.-D., and Spelke, E. S. (2009). Spontaneous mapping of number and space in adults and young children. *Cognition* 110, 198–207. doi: 10.1016/j.cognition.2008.11.003

de Hevia, M.-D., and Spelke, E. S. (2010). Number–space mapping in human infants. *Psychol. Sci.* 21, 653–660. doi: 10.1177/0956797610366091

de Hevia, M.-D., Vanderslice, M., and Spelke, E. S. (2012). Cross-dimensinal mapping of number, length and brightness by preschool children. *PLoS ONE* 7:e35530. doi: 10.1371/journal.pone.0035530

Ebersbach, M. (in press). Evidence for a spatial-numerical association in kinder-gartners using a number line task. *J. Cogn. Dev.* doi: 10.1080/15248372.2013.805134

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., and Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5-to 9-year old children: evidence for a segmented linear model. *J. Exp. Child Psychol.* 99, 1–17. doi: 10.1016/j.jecp.2007.08.006

Ebersbach, M., Luwel, K., and Verschaffel, L. (2013). Further evidence for a spatial-numerical association in children before formal schooling. *Exp. Psychol.* doi: 10.1027/1618-3169/a000250. [Epub ahead of print].

Fischer, U., Moeller, K., Bientzle, M., Cress, U., and Nuerk, H.-C. (2011). Sensorimotor spatial training of number magnitude representation. *Psychon. Bull. Rev.* 18, 177–183. doi: 10.3758/s13423-010-0031-3

Gebuis, T., and Gevers, W. (2011). Numerosities and space; indeed a cognitive illusion! A reply to de Hevia and Spelke (2009). *Cognition* 121, 248–252. doi: 10.1016/j.cognition.2010.09.008

Gevers, W., and Lammertyn, J. (2005). The hunt for SNARC. *Psychol. Sci.* 47, 10–21.

Gevers, W., Santens, S., Dhooge, E., Chen, Q., Van den Bossche, L., Fias, W., et al. (2010). Verbal-spatial and visuospatial coding of number–space interactions. *J. Exp. Psychol. Gen.* 139, 180. doi: 10.1037/a0017688

Hoffmann, D., Hornung, C., Martin, R., and Schiltz, C. (2013). Developing number–space associations: SNARC effects using a color discrimination task in 5-year-olds. *J. Exp. Child Psychol.* 116, 775–791. doi: 10.1016/j.jecp.2013.07.013

Hubbard, E. M., Piazza, M., Pinel, P., and Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nat. Rev. Neurosci.* 6, 435–448. doi: 10.1038/nrn1684

Cohen Kadosh, R., Cohen Kadosh, K., Kaas, A., Henik, A., and Goebel, R. (2007). Notation-dependent and-independent representations of numbers in the parietal lobes. *Neuron* 53, 307–314. doi: 10.1016/j.neuron.2006.12.025

Le Corre, M., and Carey, S. (2007). One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition* 105, 395–438. doi: 10.1016/j.cognition.2006.10.005

Libertus, M. E., Starr, A., and Brannon, E. M. (2014). Number trumps area for 7-month-old infants. *Dev. Psychol.* 50, 108–112. doi: 10.1037/a0032986

Lourenco, S. F., and Longo, M. R. (2010). General magnitude representation in human infants. *Psychol. Sci.* 21, 873–881. doi: 10.1177/0956797610370158

Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2009). Children's early mental number line: Logarithmic or decomposed linear? *J. Exp. Child Psychol.* 103, 503. doi: 10.1016/j.jecp.2009.02.006

Muldoon, K., Simms, V., Towse, J., Menzies, V., and Yue, G. (2011). Cross-cultural comparisons of 5-year-olds' estimating and mathematical ability. *J. Cross Cult. Psychol.* 42, 669–681. doi: 10.1177/0022022111406035

Nuerk, H. C., Iversen, W., and Willmes, K. (2004). Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *Q. J. Exp. Psychol. Sec. A* 57, 835–863. doi: 10.1080/02724980343000512

Opfer, J. E., and Furlong, E. E. (2011). How numbers bias preschoolers' spatial search. *J. Cross Cult. Psychol.* 42, 682–695. doi: 10.1177/0022022111406098

Opfer, J. E., and Thompson, C. A. (2006). "Even early representations of numerical magnitude are spatially organized: Evidence for a directional magnitude bias in pre-reading preschoolers," in *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, eds R. Sun and N. Miyake (Mahwah, NJ: Erlbaum), 639–644.

Opfer, J. E., Thompson, C. A., and Furlong, E. E. (2010). Early development of spatial–numeric associations: evidence from spatial and quantitative performance of preschoolers. *Dev. Sci.* 13, 761–771. doi: 10.1111/j.1467-7687.2009.00934.x

Patro, K., and Haman, M. (2012). The spatial-numerical congruity effect in preschoolers. *J. Exp. Child Psychol.* 111, 534–542. doi: 10.1016/j.jecp.2011.09.006

Proctor, R. W., and Cho, Y. S. (2006). Polarity correspondence: a general principle for performance of speeded binary classification tasks. *Psychol. Bull.* 132, 416. doi: 10.1037/0033-2909.132.3.416

Shaki, S., and Fischer, M. H. (2008). Reading space into numbers - a cross-linguistic comparison of the SNARC effect. *Cognition* 108, 590–599. doi: 10.1016/j.cognition.2008.04.001

Shaki, S., Fischer, M. H., and Göbel, S. M. (2012). Direction counts: a comparative study of spatially directional counting biases in cultures with different reading directions. *J. Exp. Child Psychol.* 112, 275–281. doi: 10.1016/j.jecp.2011.12.005

Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi: 10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–250. doi: 10.1111/1467-9280.02438

Slusser, E., Santiago, R., and Barth, H. (2013). Developmental change in numerical estimation. *J Exp. Psychol. Gen.* 142, 193–208. doi: 10.1037/a0028560

Sury, D., and Rubinsten, O. (2012). "Ordinal processing of numerical and non-numerical information," in *Reading, Writing, Mathematics and the Developing Brain: Listening to Many Voices*, eds Z. Breznitz, O. Rubinsten, V. J. Molfese, and D. L. Molfese (Dordrecht: Springer), 209–232. doi: 10.1007/978-94-007-4086-0_13

Szűcs, D., Nobes, A., Devine, A., Gabriel, F. C., and Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Front. Psychol.* 4:444. doi: 10.3389/fpsyg.2013.00444

Turconi, E., Jemel, B., Rossion, B., and Seron, X. (2004). Electrophysiological evidence for differential processing of numerical quantity and order in humans. *Cogn. Brain Res.* 21, 22–38. doi: 10.1016/j.cogbrainres.2004.05.003

Tversky, B., Kugelmass, S., and Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cogn. Psychol.* 23, 515–557. doi: 10.1016/0010-0285(91)90005-9

Walsh, V. A. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* 7, 483–488. doi: 10.1016/j.tics.2003.09.002

Wood, G., Willmes, K., Nuerk, H. C., and Fischer, M. H. (2008). On the cognitive link between space and number: a meta-analysis of the SNARC effect. *Psychol. Sci. Q.* 50, 489–525.

Wynn, K. (1990). Children's understanding of counting. *Cognition* 36, 155–193. doi: 10.1016/0010-0277(90)90003-3

Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing: The SNARC effect and the reverse SNARC effect in English and Arabic monoliterates, biliterates, and illiterate Arabic speakers. *J. Cogn. Cult.* 5, 165–190. doi: 10.1163/1568537054068660

The importance of being relevant: modulation of magnitude representations

Tali Leibovich[1,2]*, Liana Diesendruck[3], Orly Rubinsten[4] and Avishai Henik[2]

[1] The Cognitive Neuropsychology Laboratory, Department of Cognitive Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[2] The Cognitive Neuropsychology Laboratory, Department of Psychology and the Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[3] Department of Computer Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel
[4] Department of Learning Disabilities, University of Haifa, Haifa, Israel

The current study aims to answer two main questions. First, is there a difference between the representations of the numerical and the physical properties of visually presented numbers? Second, can the relevancy of the dimension change its representation? In a numerical Stroop task, participants were asked to indicate either the physically or the numerically larger value of two digits. The ratio between the physical sizes and the numerical values changed orthogonally from 0.1 (the largest difference) to 0.8. Reaction times (RT) were plotted as a function of both physical and numerical ratios. Trend analysis revealed that while the numerical dimension followed Weber's law regardless of task demands, the physical ratio deviated from linearity. Our results suggest that discrete and continuous magnitudes are represented by different yet interactive systems rather than by a shared representation.

Keywords: numerical cognition, ratio, Weber's law, comparative judgment, numerical Stroop

INTRODUCTION

By a very early age, we estimate and compare discrete (number of items in a group) and continuous (brightness, loudness, size, etc.) magnitudes. This ability is important for survival across species. Honeybees distinguish between flowers by "counting" the numbers of petals (Leppik, 1956), lions assess the number of their opponents by listening to their roars, and act accordingly in order to survive (McComb et al., 1994). There are many other examples [e.g., fish (Agrillo et al., 2008); birds (Koehler, 1951); avian (Lyon, 2003); amphibians (Uller et al., 2003)].

How different magnitudes are represented and processed is one of the pressing questions in numerical cognition literature. A recent theory—the *approximate number system* (ANS) (Cantlon et al., 2009)—emphasizes the commonalities between discrete and continuous magnitudes (numbers, numerosity, time, physical size, brightness, etc.) and suggests that all these magnitudes are processed by a common algorithm. One of the hallmarks of the ANS is that performance in comparative judgments of the investigated magnitudes is best described by Weber's law. Namely, the ability to discriminate between two magnitudes depends on the ratio between them (the ratio effect). This ratio dependency is also the distinguishing feature of *core system 1* suggested by Feigenson et al. (2004). *Core system 1* is a system that represents approximate numerical magnitudes independently from non-numerical properties. Note, however, that *core system 1* relates only to non-symbolic quantities. Walsh (2003) suggested that all magnitudes, in all modalities, are represented in the same region in the brain—the parietal lobe. Moreover, he argues that since they are all needed to allow us to physically interact with the environment, the purpose of all magnitude processing is to guide motor actions.

Numbers are special magnitudes; they are symbols of size. No other species has developed this kind of representation. Dehaene and Akhavein (1995) suggested that numbers are a type of special language. According to this view, we share with other animals the "number sense"—a system that enables us to crudely estimate quantities. Alongside the "number sense," which allows for an approximate representation of numbers, exists a symbolic representation in the form of numbers that enables us to accurately represent magnitudes. Several findings support that claim. Studies conducted with adults from secluded Amazonian tribes with a very limited numerical lexicon revealed that the representation of magnitudes is less accurate than that of participants from the West of the same age (Gordon, 2004; Pica et al., 2004). However, other cultural differences may contribute to these differences; in cultures with a limited numerical lexicon, children are not taught mathematics systematically—for example, they are not exposed to linear representations of numbers (such as on a ruler).

Izard and Dehaene's (2008) study provides additional evidence that the semantic meaning of numbers contributes to an exact representation of magnitudes. In this study participants were briefly exposed to arrays of dots and had to estimate their numerosity. One group of participants was first introduced to a standard array and told that it was made up of 30 dots. Magnitude estimations made by this group were more accurate than the group that was not introduced to a standard. The authors concluded that the verbal numerical value given for the standard

"calibrated" the mental number line. In order for that manipulation to take effect, one must understand the meaning of the number. For example, introducing a standard of 500 dots to a child that can count only up to 100 does not create the same calibration effect.

In light of these differences, Cohen Kadosh et al. (2005) reviewed studies that searched for the commonalities and differences between numbers and other magnitudes, such as physical size, time, and brightness. Behaviorally, a representation is considered to be shared among different magnitudes if these magnitudes are characterized by similar effects, specifically, the distance and size effects. The distance effect was first discovered for numbers by Moyer and Landauer (1967). The authors presented adult participants with pairs of numbers, different numerical distances apart, and asked them to indicate the numerically larger number. The numerical distance was found to modulate reaction times (RT). Namely, RT was faster as the distance between the numbers grew (e.g., faster response for [2 8] than for [6 8])—the distance effect. In addition, the size of the numbers affected RT; for a constant distance, RT for two small numbers (e.g., 2 3) was faster than for two large numbers (e.g., 7 8)—the size effect. Note that while the distance effect alone explains a significant part of the variance, the ratio (i.e., smaller divided by larger magnitude) explains more variance in number comparisons (Moyer and Landauer, 1967). Similar effects were found for comparison of physical sizes such as line length, brightness, and angles (Cohen Kadosh et al., 2005).

Another way to approach the question of shared representations is by looking at interactions when comparing two different dimensions. In a study by Henik and Tzelgov (1982), participants were presented with two digits and were asked to indicate the larger number (with respect to either physical size or numerical value). In congruent trials, physically larger numbers were also numerically larger (e.g., 4 2). In incongruent trials, physically larger numbers were numerically smaller (e.g., 4 2). In neutral trials, only one dimension was manipulated. In a physical task, neutral trials included the same number in two different physical sizes (e.g., 2 2), and in a numerical task, neutral trials included two different numbers of the same physical size (e.g., 2 4). Responses were influenced by the degree of congruency between relevant and irrelevant dimensions (congruent trials were faster than neutrals, and incongruent trials were the slowest), suggesting automatic processing of numerical values, even when these values were irrelevant to the task: the size-congruity effect. Moreover, manipulating the numerical distance had an effect even when it was irrelevant to the task. Hence, numerical distances were automatically computed in this task and they affected relevant judgment. The size-congruity effect was found with other continuous magnitudes such as brightness (Cohen Kadosh et al., 2008a) and the height of the number (Rubinsten and Henik, 2005). According to Cohen Kadosh et al. (2005), the size-congruity effect "suggests that different types of magnitude tap the same magnitude mechanism" (p. 1283).

To summarize, in the current literature, the presence of size and distance effects and the size-congruity effect in different magnitudes is brought as evidence of a shared representation. We propose that although comparative judgment of different dimensions results in a ratio effect, there might be subtler differences indicating that these magnitudes are processed by different systems. We shall now explain that suggestion.

Some studies in numerical cognition literature have concluded that ratio dependency is evidence of compliance with Weber's law (see Odic et al., 2013). However, Weber's law is concerned with *linear ratio dependency*. Thus, performance for a specific magnitude might be described as ratio-dependent but non-linear. Why is that distinction important? Plotting RT as a function of the ratio between two magnitudes [smaller divided by larger; see Cantlon and Brannon (2006)] suggests that for a fixed-size ratio increment of X, RT will increase by a constant amount. Thus, the difference between responses to ratios 0.2 and 0.3 is identical to the difference between responses to ratios 0.7 and 0.8. Namely, discriminability increases linearly, yielding a linear trend. In contrast, a non-linear ratio dependency, which can be described by a power function (e.g., $Y = ax^b + c$ with $b > 1$), would suggest that RT does not change by a constant amount. RT increases, non-linearly, with the similarity (of size, for example) between the stimuli. For example, an increment from ratio 0.2 to 0.3 produces a lower increase in RT than an increment in size ratio from 0.7 to 0.8. This would imply that discriminability becomes more difficult with increase in similarity.

THE CURRENT STUDY

In the current study, participants were presented with two numbers and were asked to choose either the physically or the numerically larger number. The ratio between the magnitudes (smaller divided by larger magnitude; for physical size or numerical value) varied from 0.1 to 0.8, with 0.1 being the largest difference (e.g., 2 and 7 or 2 2), and 0.8 being the smallest difference (e.g., 6 and 8 or 2 2). Next, we plotted RT as a function of the magnitude ratio for every task. In that way, we could examine more closely possible differences between performance in comparison of numerical values and physical sizes. In Experiment 1, only one dimension was manipulated [i.e., the neutral condition in a numerical Stroop task as described in Henik and Tzelgov (1982)]. For every participant and for every task, RT was plotted as a function of the ratio between the to-be-compared magnitudes. These plots were compared once to a power function ($RT = ax^b + c$) and once to a linear function, and the fit values (r^2) were recorded. If the fit to the power function when $b = 1$ and the fit to the linear function did not differ, it indicated that the trend was linear. However, if the fit to the power function when $b \neq 1$ was better than the fit to the linear function, then the trend deviated from linearity. Note that the exponent is not just one more free parameter that can explain more variance, because if the trend is linear, the exponent $b = 1$ would force the power function to be linear as well. Thus, the exponent can only add to the explained variance if different than 1.

Based on previous findings, we expected that performance in the numerical task would be best described by Weber's law; first, Weber's law has been suggested in studies that examined numerical discrimination (Moyer and Landauer, 1967; Cantlon and Brannon, 2006), and second, as suggested by Izard and Dehaene (2008), the exact numerical values allow us to represent magnitudes more accurately. There is no dispute about the value of the

number 3, or that the difference between 9 and 4 is exactly 5. This is true for all the number pairs used here (i.e., numbers from 1 to 9). Thus, it is reasonable to believe that the level of difficulty will increase with increase in similarity between the stimuli[1].

Unlike the exact verbal representation of the difference between two numbers (i.e., 3 and 5), we cannot tell the exact *physical* size of the number 3 and we cannot tell the exact difference in size between 3 and 3. For that reason, we might expect deviations from Weber's law. In addition, representation of the mental number line was found to be modulated by other factors, such as attention. For example, Anobile et al. (2011) found that under attentionally demanding conditions, an otherwise linear mapping becomes compressed and non-linear. It is possible that the verbal aspect of numbers has a similar effect on the representation of their physical size. Accordingly, for the physical task, we asked whether performance in the physical comparison task would comply with Weber's law or not.

In Experiment 2, both numerical and physical dimensions were orthogonally manipulated to create congruent and incongruent conditions [as described in Henik and Tzelgov (1982)]. RT was plotted as a function of both the physical and the numerical magnitudes. Here we investigated if (and how) the relevancy of the dimension influences performance. Assuming that physical and numerical magnitudes are processed by different systems/mechanisms and that numerical values are automatically processed, we would expect that: (1) performance in the numerical dimension would comply with Weber's law regardless of task demands; (2) when the physical dimension is relevant, performance in the physical task might deviate from Weber's law; and (3) when the physical dimension is irrelevant (in the numerical task), performance in the physical dimension might not deviate from Weber's law due to an interaction with a very accurate numerical magnitude representation.

Questions regarding shared or separate representation of different magnitudes and the interactions between such representations are of great developmental interest. Odic et al. (2013) had 3- to 6-year-olds discriminate between quantities (numbers) or the area of irregular shapes. They found that the acuity of area discrimination was better than number discrimination. Namely, participants were able to detect much smaller changes in area than in numerosity. However, both area acuity and number acuity showed a similar growth function throughout development. Thus, the authors raised the possibility of similar yet separated development for different magnitudes. Similarly, Lourenco and Longo (2010) suggested that in early age, all magnitudes are represented by a shared system. However, with time, there is a division of this system into subsystems that specialize in the processing of specific magnitudes. Some theories suggest that understanding non-symbolic number is the basis for understanding symbolic numbers. Von Aster and Shalev (2007), for example, suggested a 4-step developmental model of numerical cognition. According to this model, we are born with the ability to represent and approximate the cardinality of magnitudes. This ability provides the basic meaning of numbers (step 1). In step 2, the child learns to associate quantity with number words, and in step 3, with the Arabic numeral symbols. The association to Arabic symbols is a precondition for the development of the mental number line (step 4). In step 4, ordinality is represented as a second (and acquired) core system for numbers.

The current experiment provides a picture about the representation of symbolic (numbers) and non-symbolic (physical size) magnitudes (Experiment 1), and the interaction between these representations (Experiment 2) in adults. This data and this methodology can serve as a baseline for studies aiming to investigate representation of symbolic and non-symbolic sizes (and an interaction between them) at different points in normal and impaired development.

EXPERIMENT 1—UNI-DIMENSIONAL COMPARISONS
METHODS
Participants
Fourteen volunteers (11 females, 3 male, mean age: 23 years), students from Ben-Gurion University of the Negev or Achva Academic College, participated in the experiment for class credit. All participants had intact or corrected vision and no learning disabilities. Seven performed the physical task first and seven performed the numerical task first.

Stimuli
Each stimulus was composed of two digits from 1 to 9. The numbers (Courier New font) appeared in lime color on a black background, each 1.75 cm from the center of a computer screen (i.e., center of the number to center of the screen). The participants sat at a distance of about 50 cm from the screen. Numbers (1–9) were paired to create eight numerical ratios (0.1–0.8). The ratios were rounded to one digit after the decimal point (see **Table 1**). Note that the ratio is a continuum. Thus, although the ratios in every category may differ from pair to pair, all the pairs with ratio 0.3, for example, are larger than 0.2 and smaller than 0.4 [see similar design in Cantlon and Brannon (2006)]. For example, to create the numerical ratio of 0.3, we used the pair 2 and 6 (2/6 = 0.3). Similarly, nine font sizes (12.5, 25, 37.5, 50, 62.5, 75, 87.5, 100, and 112.5) were paired in order to create eight physical size ratios (0.1–0.8). The physical sizes were adopted from Cohen Kadosh et al. (2008b)—Experiment 2. The pairs of physical sizes we used are outlined in **Table 2**. For example, to create the physical ratio of 0.5, we used the fonts 50 and 100 (50/100 = 0.5) or 25 and 62.5, or 37.5 and 75. The same sizes were used for more than one physical ratio to avoid confounding of size and ratio.

Experiment 1 included three physical blocks and three numerical blocks. In the physical block, the same number appeared twice in different physical sizes. In total, a physical block contained 96 stimuli: 8 physical ratios (0.1–0.8) × 2 sides (larger

---

[1]"Exact representation" refers to a specific value that goes with the numerical symbol (i.e., 3 represents exactly 3 items). This is in contrast to physical sizes, such as length, that do not have verbal labels to represent them, and which we can only estimate their exact size. Both exact and non-exact magnitudes can be placed on a magnitude line, which is logarithmic with overlapping representations. However, as suggested by Izard and Dehaene (2008), being able to represent exact and not estimated size, might "calibrate" the mental number line. Namely, we still expect a ratio effect but this ratio effect might have different manifestations (such as different exponents).

Table 1 | Pairs of stimuli by numerical ratio.

Category ratio	Ratio	Large number	Small number
0.1	0.11	9	1
	0.13	8	1
	0.14	7	1
0.2	0.20	5	1
	0.22	9	2
	0.25	4	1
	0.25	8	2
0.3	0.33	3	1
	0.33	6	2
	0.33	9	3
0.4	0.40	5	2
	0.43	7	3
	0.44	9	4
0.5	0.50	2	1
	0.50	4	2
	0.50	6	3
	0.50	8	4
0.6	0.60	5	3
	0.63	8	5
	0.67	3	2
	0.67	6	4
	0.67	9	6
0.7	0.71	7	5
	0.75	4	3
	0.75	8	6
0.8	0.80	5	4
	0.83	6	5
	0.86	7	6

Ratio, (small number/large number) with an accuracy of 2 decimal places.

Table 2 | Pairs of stimuli by physical ratio.

Category ratio	Ratio	Large font size	Small font size
0.1	0.11	112.5	12.5
	0.13	100	12.5
	0.14	87.5	12.5
0.2	0.20	62.5	12.5
	0.22	112.5	25
	0.25	50	12.5
	0.25	100	25
0.3	0.33	37.5	12.5
	0.33	75	25
	0.33	112.5	37.5
0.4	0.40	62.5	25
	0.43	87.5	37.5
	0.44	112.5	50
0.5	0.50	25	12.5
	0.50	50	25
	0.50	75	37.5
	0.50	100	62.5
0.6	0.60	62.5	37.5
	0.63	100	62.5
	0.67	37.5	25
	0.67	75	50
	0.67	112.5	75
0.7	0.71	87.5	62.5
	0.75	50	37.5
	0.75	100	75
0.8	0.80	62.5	50
	0.83	75	62.5
	0.86	87.5	75

Ratio, (small size/large size) with an accuracy of 2 decimal places.

number on left vs. on right) × 6 pairs of numbers. In the numerical block, different numbers appeared in the same physical size. In total, a numerical block contained 96 stimuli: 8 numerical ratios (0.1–0.8) × 2 sides (larger number on left vs. on right) × 6 pairs of numbers. In both tasks, the specific pairs of numbers and their specific physical sizes within a given ratio were randomly selected for every participant.

Procedure
Participants were asked to decide, as quickly as possible while avoiding errors, which of the two numbers was physically larger (in the physical block), or numerically larger (in the numerical block). They were asked to indicate their decision by pressing a key (*p* or *q*) corresponding to the side of the larger number. Each trial began with a central fixation point presented for 300 ms. Five hundred ms after the elimination of the fixation point, a pair of numbers appeared and remained in view until the participant pressed a key. The next trial started 500 ms after response onset
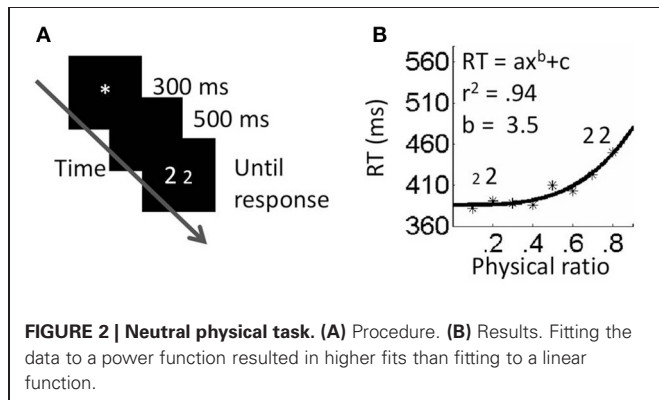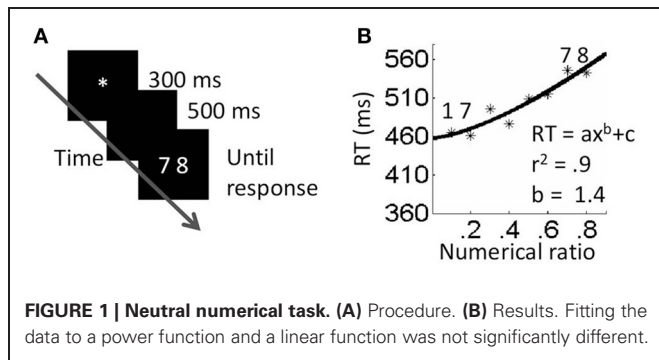
(see **Figures 1A**, **2A**). For every task, instructions and six practice trials were presented first, followed by three experimental blocks. The stimuli within a block appeared in a random order.

Design
For each task there was a single independent variable of ratio that had 8 possible values. The dependent measures were RT and accuracy.

RESULTS
We calculated error rates and mean RT in milliseconds (ms) for correct responses only, for every ratio (numerical and physical). Very low (less than 150 ms) and very high (over 3000 ms) RTs were excluded from the analysis (1 trial in the numerical task). These mean RTs were subjected to a One-Way analysis of variance (ANOVA) with ratio as an independent variable. The main effects of numerical and physical ratios were significant [$(F_{(7, 91)} = 32.49, MSE = 465, p < 0.001, \eta_p^2 = 0.71)$ and $(F_{(7, 91)} = 18.92,$

FIGURE 1 | Neutral numerical task. (A) Procedure. **(B)** Results. Fitting the data to a power function and a linear function was not significantly different.

FIGURE 2 | Neutral physical task. (A) Procedure. **(B)** Results. Fitting the data to a power function resulted in higher fits than fitting to a linear function.

$MSE = 397$, $p < 0.001$, $\eta_p^2 = 0.59$), respectively]. Namely, RT increased with magnitude ratio. Accuracy for the numerical task (mean $= 0.97$, $SD = 0.16$) and for the physical task (mean $= 0.98$, $SD = 0.14$) presented a similar pattern: for the numerical task: $F_{(7,91)} = 25.7$, $MSE = 0.0003$, $p < 0.001$, $\eta_p^2 = 0.66$, and for the physical task: $F_{(7, 91)} = 13.84$, $MSE = 0.0004$, $p < 0.001$, $\eta_p^2 = 0.52$.

Exponents analyses

In order to investigate whether numerical and physical comparisons result in different functions, we plotted each participant's RT as a function of the magnitude ratio (i.e., physical ratio or numerical ratio) and fitted the plots to a power function ($RT = ax^b + c$) using the Matlab curve fitting tool. Then, the exponent values were used as a dependent variable in a t-test for dependent samples. Three participants were removed from this analysis due to r^2 that deviated from the average by more than two statistical deviations. According to the results, the difference between the exponents for physical comparisons and numerical comparisons was significant. Specifically, the exponent was higher for the physical task (average exponent $= 3.46$, $SD = 2.2$) than for the numerical task (average exponent $= 1.44$, $SD = 0.75$) $t_{(10)} = 2.92$, $p < 0.05$. A one-sample t-test revealed that while the exponents in the physical task were significantly different than one [$t_{(10)} = 3.4$, $p < 0.01$], the exponents in the numerical task were not significantly different than one [$t_{(10)} = 1.71$, ns]. This suggests that performance in the physical comparison task, but not the numerical comparison task, deviated from linearity, thus violating Weber's law.

To further strengthen this suggestion, we fitted the data of every participant in every task twice; once to a linear function and once to a power function. We then used the fit values (r^2) as the dependent measures in t-tests for dependent samples. This was done for each task separately. For the numerical task, there was no significant difference between the fits of the two functions ($t < 1$, ns). This was expected since in this task the exponent values calculated by the fitting process were close to one; this means that, in practice, the exponential equation behaved as a linear one. For the physical task, r^2 values were higher when the data was fitted to a power function than when fitted to a linear function $t_{(12)} = 2.9$, $p = 0.01$.

DISCUSSION

The results of Experiment 1 revealed a different relationship between the RT and magnitude ratio for physical sizes and numerical values and suggest that the two dimensions have different representations. In the numerical task, performance complied with Weber's law. Namely, discriminability increased linearly. In contrast, in the physical task RT did not change by a constant amount, violating Weber's law; rather, for a fixed increment in size ratio, RT increased with the similarity between the stimuli, although not linearly.

We suggest that the difference between the representations (numerical and physical properties) might stem from the different nature of the stimuli: numbers are a special kind of magnitude: they are discrete, countable, symbolic representations with verbal labels. Physical sizes, on the other hand, are non-countable, continuous magnitudes that one can only estimate. Thus, while numbers are represented on a "mental *number* line" that complies with Weber's law, physical sizes may be represented on a more general "mental *magnitude* line" that is noisier due to the nature of continuous properties. Our data cannot determine between the possibility of one mental magnitude line with different levels of noise for different representations, or two separate systems: one that represents numbers, and one that represents continuous properties [similar to the suggestions of Odic et al. (2013) and Lourenco and Longo (2010)].

The current study involved adults. It will be interesting to use this design with children who are just starting to learn the numerical symbols system; if our hypothesis is correct and the linear representation is due to an exact verbal representation of numbers and the difference between them, then what trend will children early in their formal education produce? Studying this trend, and not only the existence of a ratio effect, can be more informative and detect more subtle changes in performance.

EXPERIMENT 2—NUMERICAL STROOP TASKS

Given the congruity effect and the difference between the representations of symbolic and non-symbolic dimensions observed in Experiment 1, it is interesting to ask what happens to these representations in a numerical Stroop task, like the one employed by Henik and Tzelgov (1982). Can the relevancy of a dimension modulate mental representations? Can different representations co-exist? The following experiment addresses these questions.

METHODS

Participants

Twenty volunteers (15 females, 5 males, mean age: 22.95 years), students from Ben-Gurion University of the Negev or Achva Academic College, participated in the experiment for class credit. All participants had intact or corrected vision, and no learning disabilities. Ten participants performed the physical task and 10 performed the numerical task.

Stimuli

The same physical sizes and numerical ratios of digits from Experiment 1 were used here. The stimuli created two congruency conditions: congruent or incongruent, as described by Henik and Tzelgov (1982) (see **Figure 3**). Similar to Experiment 1, instead of manipulating the distance between the numbers, we manipulated the ratio of both dimensions. We used eight physical ratios and eight numerical ratios. An experimental block (numerical block as well as physical block) contained 256 stimuli: 2 conditions (congruent, incongruent) × 8 physical ratios (0.1–0.8) × 8 numerical ratios (0.1–0.8) × 2 sides of presentation. The block repeated 14 times, 7 times in a session. In every block, the specific numbers and their absolute size were randomly selected.

Procedure

The procedure was similar to that of Experiment 1 except that blocks included congruent and incongruent trials, as can be seen in **Figure 3**. The experiment was completed in two sessions.

RESULTS

In the physical task, the average accuracy rate was 0.97 ($SD = 0.17$), and in the numerical task, the average accuracy rate was 0.95 ($SD = 0.21$). In both tasks, there was not enough variance to analyze accuracy rates for the different conditions.

Mean RT in milliseconds was calculated for correct responses only. Very high (over 3000 ms) and very low (under 150 ms) RTs were eliminated from the analysis (3 trials—only in the numerical task). Mean RTs were subjected to a three-way ANOVA with physical size ratio (0.1–0.8), numerical size ratio (0.1–0.8) and

FIGURE 3 | Procedure of Experiment 2. Trials in these blocks were either congruent or incongruent.

congruity (congruent and incongruent) as independent variables. Physical and numerical tasks were analyzed separately.

In the physical task, the three main effects were significant: physical ratio, $F_{(7, 63)} = 103.12$, $MSE = 4472$, $p < 0.001$, $\eta_p^2 = 0.92$; numerical ratio, $F_{(7, 63)} = 4.04$, $MSE = 398$, $p < 0.001$, $\eta_p^2 = 0.31$; and congruity, $F_{(1, 9)} = 64.88$, $MSE = 2560$, $p < 0.001$, $\eta_p^2 = 0.88$. The effect of the physical and the numerical ratios can be seen in **Figures 4A,B**; RT was the slowest when the ratio between the physical magnitudes (the relevant dimension) was 0.8 (the smallest difference), and the ratio between the numerical values (the irrelevant dimension) was 0.1 (the largest difference). Physical ratio was found to influence congruity [$F_{(7, 63)} = 18.8$, $MSE = 1374$, $p < 0.001$, $\eta_p^2 = 0.68$] as did numerical ratio [$F_{(7, 63)} = 6.4$, $MSE = 729$, $p < 0.001$, $\eta_p^2 = 0.42$]. As can be seen in **Figure 4C**, the congruity effect was the strongest when the ratio between the physical magnitudes was 0.8, and the ratio between the numerical values was 0.1. The triple interaction between physical ratio, numerical ratio and congruity was not significant [$F_{(49, 441)} = 1.2$, ns, $\eta_p^2 = 0.12$].

In the numerical task, the three main effects were significant: physical ratio, $F_{(7, 63)} = 67.08$, $MSE = 1250$, $p < 0.001$, $\eta_p^2 = 0.88$; numerical ratio, $F_{(7,63)} = 113.15$, $MSE = 977$, $p < 0.001$, $\eta_p^2 = 0.93$; and congruity, $F_{(1, 9)} = 78.11$, $MSE = 8795$, $p < 0.001$, $\eta_p^2 = 0.9$. The effect of the physical and the numerical ratios can be seen in **Figures 5A,B**; RT was the slowest when the ratio between the numerical values (the relevant dimension) was 0.8 (the smallest difference), and the ratio between the physical sizes (the irrelevant dimension) was 0.1 (the largest difference). Physical ratio was found to influence congruity [$F_{(7, 63)} = 19.61$, $MSE = 808$, $p < 0.001$, $\eta_p^2 = 0.69$] as did numerical ratio [$F_{(7, 63)} = 13.43$, $MSE = 410$, $p < 0.001$, $\eta_p^2 = 0.6$]; this can be seen in **Figure 5C**. The congruity effect was strongest when the ratio between the numerical values was 0.8, and the ratio between the physical sizes was 0.1. The triple interaction between physical ratio, numerical ratio and congruity was not significant [$F_{(49, 441)} = 1.2$, ns, $\eta_p^2 = 0.12$].

### Trend analyses

Similar to the previous experiment, we fitted the data to two functions:

$$RT = ax + cy + d. \tag{1}$$

$$RT = ax^b + cy + d. \tag{2}$$

In these functions, x represents the physical dimension (e.g., physical ratios from 0.1 to 0.8) and y represents the numerical dimension (e.g., numerical ratios from 0.1 to 0.8); d indicates the minimal RT; and b in function (2) is the exponent of the physical dimension. These functions are derived from the ANOVA main effects for physical (x) and numerical (y) ratios. The absence of a combined xy component in the functions reflects the lack of interaction between these two dimensions. According to our results from Experiment 1, function (1) is expected to give the best fit according to Weber's law (i.e., should fit to the results of the numerical task of Experiment 2). On the other hand, the best fit for the physical task of Experiment 2 should be function
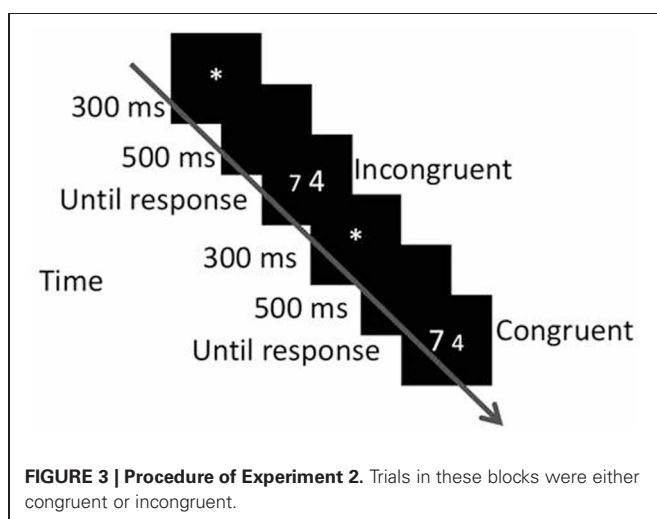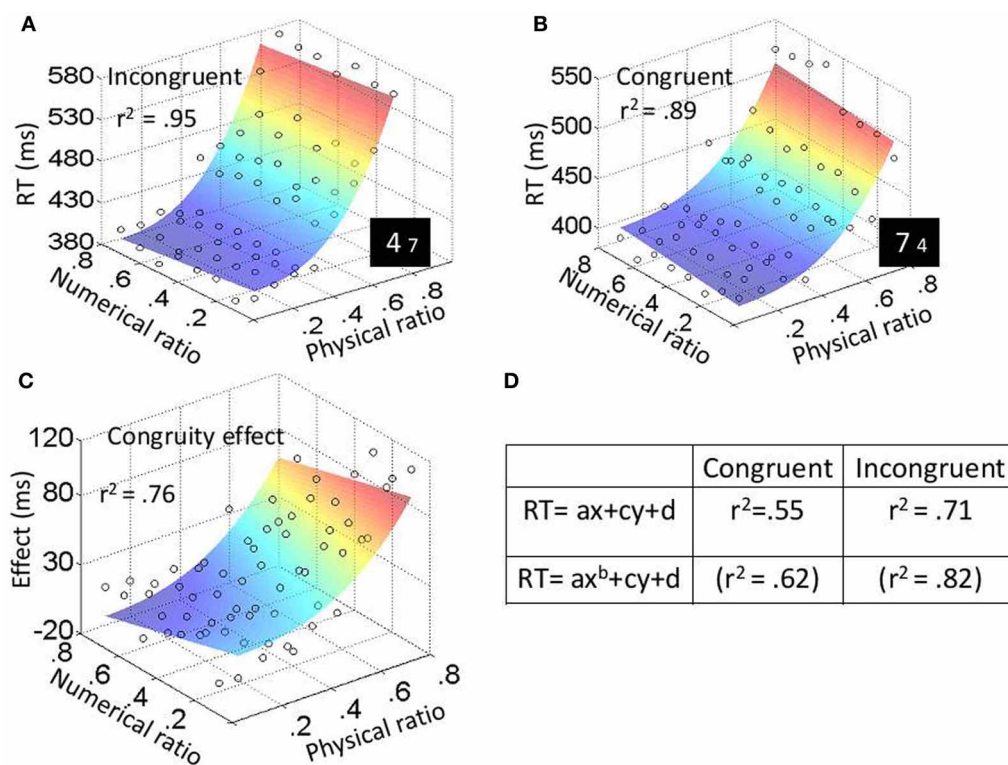
FIGURE 4 | Experiment 2: physical task results. Average RT (for every participant in every condition) was plotted as a function of both physical (x-axis) and numerical (y-axis) ratios. The color of the surface represent RTs—blue is for the lowest RTs and red for the highest RTs. Fit values for plots **(A)**, **(B)**, and **(C)** are higher than the average of the fits based on individual participants. **(D)** Fits (r^2) when fitting the plots of the different conditions to function (1) or function (2).

(2) since the numerical dimension was linear while the physical dimension deviated from linearity. Similar to Experiment 1, we fitted the plot for every participant in every task (physical or numerical) and every condition (congruent and incongruent) and recorded the fit (r^2) values (see **Figures 4D**, **5D**). These fits were then used as dependent variables in a two-way ANOVA, with task as a between-subject variable and condition as within-subject variable. For both the physical and numerical tasks, fits were higher for function (2) [$F_{(1, 18)} = 38.1, MSE = 0.002, p < 0.001, \eta_p^2 = 0.38$]. The analysis also revealed a main effect for condition, where fits were higher for incongruent than for congruent trials [$F_{(1, 18)} = 45.6, MSE = 0.01, p < 0.001, \eta_p^2 = 0.72$]. In addition, a significant Two-Way interaction between task and condition was found; the difference between the fits for the numerical task was smaller than for the physical task [$F_{(1, 18)} = 5.97, MSE = 0.002, p < 0.05, \eta_p^2 = 0.25$].

GENERAL DISCUSSION

In the current study we took advantage of the fact that: (1) visually presented numbers have two dimensions of magnitude; and (2) that numbers are processed automatically, to answer two main questions. First, is there a difference between the representations of numerical and physical magnitudes? Second, can the relevancy of the dimension change its representation? To answer these questions, participants compared the physical size or the numerical

values of pairs of numbers. In Experiment 1, only one dimension was manipulated. In Experiment 2, both dimensions were orthogonally manipulated to create congruent and incongruent conditions.

In Experiment 1, we found that performance in the numerical task fits the notion of Weber's law, as predicted by the literature. In contrast, performance in the physical task, though still ratio-dependent, deviated from Weber's law. This was evidenced by two results. First, fits for power functions were higher than for linear functions only in the physical task. Second, when fitting the plots to an exponential function, the exponents obtained in the physical task were significantly higher than those obtained in the numerical task. This pattern of results fits our suggestion that differences exist between representations of numerical and physical magnitudes. We hypothesized that these differences could be attributed to the exact verbal representation of numerical values in comparison to the less accurate nature of physical sizes. However, more research is necessary to confirm this hypothesis. One way to further test this hypothesis is through a developmental study with children in different stages of their familiarity with the numerical symbols systems. If our hypothesis is correct, plotting RT as a function of the numerical ratio, and fitting this plot to a power function, will yield an exponent greater than 1. This exponent value will become similar to 1 when the child gains experience with the numerical symbols system.
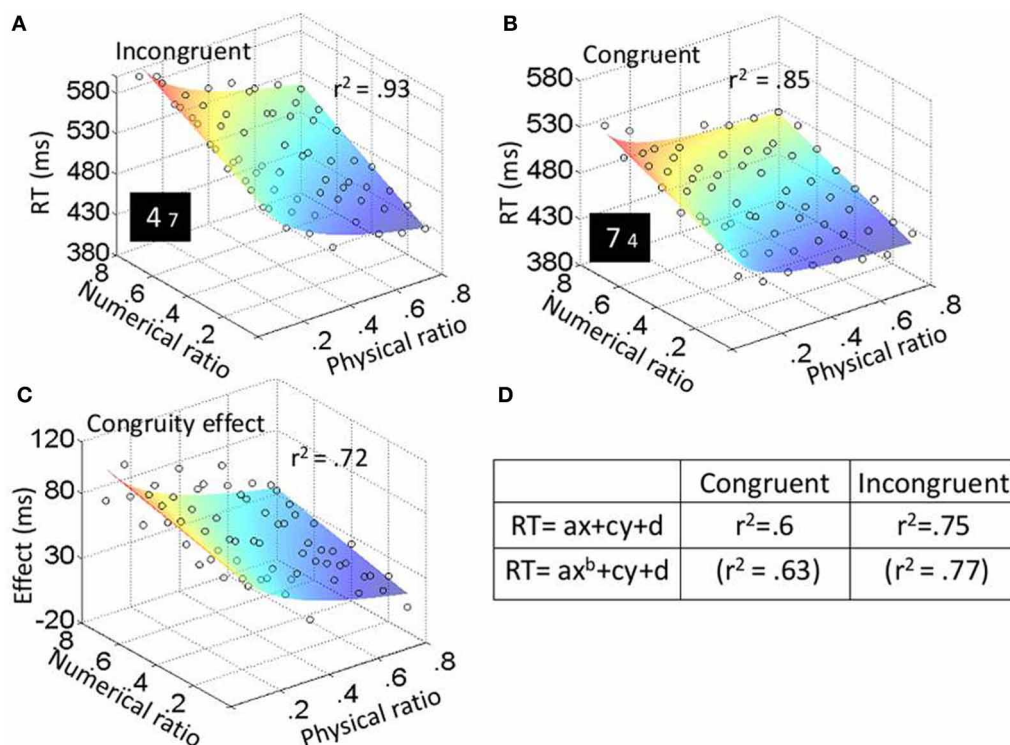
FIGURE 5 | Experiment 2: numerical task results. Average RT (for every participant in every condition) was plotted as a function of both physical (*x*-axis) and numerical (*y*-axis) ratios. The color of the surface represent RTs—blue is for the lowest RTs and red for the highest RTs. Fit values for plots **(A)**, **(B)**, and **(C)** are higher than the average of the fits based on individual participants. **(D)** Fits (*r²*) when fitting the plots of the different conditions to function (1) or function (2).

Our suggestion is in line with other findings in the literature connecting exact representation with magnitude representation. Whalen and colleagues (1999) presented participants with a target number and asked them to press a key repeatedly (without counting) until they believed that they reached the target number. These results complied with the results of a similar experiment with animals—the number of presses increased with the target number, suggesting scalar variability (i.e., encoding of magnitudes is noisy, and this noise increases proportionally with magnitude) (Gallistel and Gelman, 2000). In a similar design, participants had to reproduce time durations. The coefficient of the variance (the ratio between the variability of the estimation and its average) was higher for estimation of time duration than for non-verbal counting. Duration of time is continuous, while key pressing is discrete. This may be related to the change in the coefficient of variance in the two tasks. In a similar study, Cordes et al. (2001) conducted a key-press experiment with adults under conditions that required either counting aloud or did not allow vocal or sub-vocal counting (i.e., allowed only non-verbal counting). In the non-verbal counting condition, the authors found a power law relationship between the target number and the average number of presses, suggesting that alongside of the non-verbal counting mechanisms that we share with other animals, there exists another representation for verbal counting.

Additional evidence for the influence of semantic meaning on the representation of the mental number line comes from line-mapping experiments with children. Ebersbach et al. (2008) asked children between the ages of 5- to 9-years old to map numbers onto a number line. They found that the representation of the mental number line was influenced by familiarity; mapping was linear for familiar numbers and logarithmic for less familiar numbers. Similar results were obtained in a study by Siegler and Opfer (2003). In this study, 7-year-olds or adults had to map numbers onto a line from 0 to 100 or 0 to 1000. While both children and adults revealed the same linear mapping for 0–100. Children, who were unfamiliar with numbers above 100, mapped the numbers logarithmically when they were beyond 100.

Experiment 2 included numerical and physical Stroop tasks. Unlike previous studies, we had 8 physical and 8 numerical ratios, and plotted RT as a function of both the physical and the numerical ratio. In that way, we were able to analyze differences in the trends of the different dimensions. Our results suggested that the representation of the physical dimension depends on task demands; when the physical dimension was relevant its trend was exponential, similar to the trend of the neutral task (in Experiment 1). As a result, there was a large difference between the fits to function (1) that assumes linearity for both physical and numerical dimensions, and function (2) that allows deviation from linearity for the physical dimension. Namely, fits were

higher for function (2). In contrast, when the numerical dimension was relevant, there was a very small difference between the fits to functions (1) and (2). This suggests that the trend of the physical dimension shifted and became more "linear."

We propose that this shift of exponents is due to an interaction between the exact magnitude representations of the numerical values with the estimated magnitude representation of physical continuous dimensions. Specifically, we suggest that when participants were asked about the physical size of the numbers, they activated a spontaneous mental *magnitude* line, which is noisier and less organized than the exact number line (Izard and Dehaene, 2008). The result of such spontaneous activation of the mental *magnitude* line representation is the deviation from linearity, much as in the neutral task in Experiment 1. In contrast, when asked about the numerical value of the numbers, participants activated an exact mental number line representation. The physical sizes, in turn, could have been mapped onto that line. To confirm our suggestion, one can change the stimuli used in the experiment. For example, compare two continuous magnitudes—brightness and area of squares. According to our hypothesis, since both dimensions are continuous, their trend should be exponential, regardless of the irrelevant dimensions. Since RT was faster and accuracy was higher in the physical tasks, there is a possibility that the deviation from linearity was a result of a floor effect in small ratios, where the task was very easy to perform. This is a built-in limitation: we found here that participants were much faster in comparing sizes than comparing numbers, when comparing a wide range of ratios. This alone provides important information about the processing mechanism—something in the processing of the physical size allows it to be faster and more accurate, and to deviate from Weber's law. If we try to artificially encourage slower RT, it will no longer be comparable to the discrete task.

Our experimental design and analysis provide a tool that developmental studies in the field of numerical cognition can benefit from for several reasons. First, changes throughout development might be subtle. Using a wide range of ratios, instead of 2 or 3 ratios, and analyzing the function created when RT is plotted as a function of magnitude ratio, might uncover differences that would be missed in the commonly studied age × (2–3 points) ratio interaction. The analysis of Experiment 1 can be used to ask how different magnitudes are represented at different stages of development. The analysis used in Experiment 2—analyzing the

effect of both physical and numerical ratios on performance—can answer questions regarding an interaction between representation of numbers and physical sizes at different ages.

In conclusion, in the current work we examined the ratio effect in finer resolution compared with studies reported so far in order to detect differences between representation of numerical values and physical sizes. To the best of our knowledge, this is the first work in numerical cognition to use the exponent as a dependent variable and to investigate the combined influence of different magnitude ratios on RT in a comparative judgment task.

Our results, coupled with the current literature, suggest that numerical values and physical magnitudes have different representations. This can be the result of two different yet interacting core systems: a core system that represents continuous magnitudes, and a system that represents discrete magnitudes. These systems are shared across species. The system for continuous magnitudes is ratio-dependent but does not necessarily comply with Weber's law. Our pattern of results is in favor of a previous suggestion that the system for processing continuous magnitudes might be older than the system for processing discrete magnitudes (Cantlon et al., 2009; Henik et al., 2012), although further research is needed to support this notion. An interaction between symbolic processes (language) and a system for representation of discrete magnitudes may explain the special and exact representation of numbers, as supported by the developmental studies mentioned above. The interaction between the continuous and discrete representations is manifested in a change of trend of the physical dimension when the numerical dimension is relevant; activating the mental number line to resolve a numerical Stroop task allows for a less exponential representation of the irrelevant physical dimension. Note that the representation of continuous magnitudes on a mental magnitude line has been less investigated, and it is hard to hypothesize how some of the current models apply to continuous magnitudes. For example, Verguts et al. (2005) suggested that the representation of different quantities is described by a place-coding. It is hard to understand how the concept of continuous magnitudes can be described by a place-coding. Thus, more studies in the field are required to confirm our suggestion.

ACKNOWLEDGMENTS

REFERENCES

Agrillo, C., Dadda, M., Serena, G., and Bisazza, A. (2008). Do fish count? Spontaneous discrimination of quantity in female mosquitofish. *Anim. Cogn.* 11, 495–503. doi: 10.1007/s10071-008-0140-9

Anobile, G., Cicchini, G. M., and Burr, D. C. (2011). Linear mapping of numbers onto space requires attention. *Cognition* 122, 454–459. doi: 10.1016/j.cognition.2011.11.006

Cantlon, J. F., and Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychol. Sci.* 17, 401–406. doi: 10.1111/j.1467-9280.2006.01719.x

Cantlon, J. F., Platt, M. L., and Brannon, E. M. (2009). Beyond the number domain. *Trends Cogn. Sci.* 13, 83–91. doi: 10.1016/j.tics.2008.11.007

Cohen Kadosh, R., Cohen Kadosh, C., and Henik, A. (2008a). When brightness counts: the neuronal correlate of numerical-luminance interference. *Cereb.*

Cortex 18, 337–343. doi: 10.1093/cercor/bhm058

Cohen Kadosh, R., Henik, A., and Rubinsten, O. (2008b). Are Arabic and Verbal numbers processed in different ways? *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 1377–1391. doi: 10.1037/a0013413

Cohen Kadosh, R., Cole, N. B., Henik, A., Rubinsten, O., Mohr, H., Dori, H., et al. (2005). Are numbers special? The comparison systems of the human brain investigated by fMRI. *Neuropsychologia* 43, 1238–1248.

doi: 10.1016/j.neuropsychologia.2004.12.017

Cordes, S., Gelman, R., Gallistel, C. R., and Whalen, J. (2001). Variability signatures distinguish verbal from non-verbal counting for both large and small numbers. *Psychon. Bull. Rev.* 8, 698–707. doi: 10.3758/BF03196206

Dehaene, S., and Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *J. Exp. Psychol. Learn.*

Mem. Cogn. 21, 314–326. doi: 10.1037/0278-7393.21.2.314

Ebersbach, M., Luwel, K., Frick, A., Onghena, P., and Verschaffel, L. (2008). The relationship between the shape of the mental number line and familiarity with numbers in 5- to 9-year old children: evidence for a segmented linear model. *J. Exp. Child Psychol.* 99, 1–17. doi: 10.1016/j.jecp.2007.08.006

Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307–314. doi: 10.1016/j.tics.2004.05.002

Gallistel, C. R., and Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends Cogn. Sci.* 4, 59–65. doi: 10.1016/S1364-6613(99)01424-2

Gordon, P. (2004). Numerical cognition without words: evidence from Amazonia. *Science* 306, 496–499. doi: 10.1126/science.1094492

Henik, A., Leibovich, T., Naparstek, S., Diesendruck, L., and Rubinsten, O. (2012). Quantities, amounts, and the numerical core system. *Front. Hum. Neurosci.* 5:186. doi: 10.3389/fnhum.2011.00186

Henik, A., and Tzelgov, J. (1982). Is three greater than five: the relation between physical and semantic size in comparison tasks. *Mem. Cogn.* 10, 389–395. doi: 10.3758/BF03202431

Izard, V., and Dehaene, S. (2008). Calibrating the mental number line. *Cognition* 106, 1221–1247. doi: 10.1016/j.cognition.2007.06.004

Koehler, O. (1951). The ability of birds to count. *Bullet. Anim. Behav.* 9, 41–45.

Leppik, E. (1956). The form and function of numeral patterns in flowers. *Am. J. Bot.* 43, 445–453. doi: 10.2307/2438881

Lourenco, S. F., and Longo, M. R. (2010). General magnitude representation in human infants. *Psychol. Sci.* 21, 873–881. doi: 10.1177/0956797610370158

Lyon, B. E. (2003). Egg recognition and counting reduce costs of avian conspecific brood parasitism. *Nature* 422, 495–499. doi: 10.1038/nature01505

McComb, K., Packer, C., and Pusey, A. (1994). Roaring and numerical assessment in contests between groups of female lions, Panthera leo. *Anim. Behav.* 47, 379–387. doi: 10.1006/anbe.1994.1052

Moyer, R. S., and Landauer, T. K. (1967). Time required for judgments of numerical inequality. *Nature* 215, 1519–1520. doi: 10.1038/2151519a0

Odic, D., Libertus, M. E., Feigenson, L., and Halberda, J. (2013). Developmental change in the acuity of approximate number and area representations. *Dev. Psychol.* 49, 1103–1112. doi: 10.1037/a0029472.

Pica, P., Lemer, C., Izard, V., and Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306, 499–503. doi: 10.1126/science.1102085

Rubinsten, O., and Henik, A. (2005). Automatic activation of internal magnitudes: a study of developmental dyscalculia. *Neuropsychology* 19, 641–648. doi: 10.1037/0894-4105.19.5.641

Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation. *Psyc. Sci.* 14, 237–243. doi: 10.1111/1467-9280.02438

Uller, C., Jaeger, R., Guidry, G., and Martin, C. (2003). Salamanders (*Plethodon cinereus*) go for more: rudiments of number in an amphibian. *Anim. Cogn.* 6, 105–112. doi: 10.1007/s10071-003-0167-x

Verguts, T., Fias, W., and Stevens, M. (2005). A model of exact small-number representation. *Psychon. Bull. Rev.* 12, 66–80. doi: 10.3758/BF03196349

Von Aster, M. G., and Shalev, R. S. (2007). Number development and developmental dyscalculia. *Dev. Med. Child Neurol.* 49, 868–873. doi: 10.1111/j.1469-8749.2007.00868.x

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* 7, 483–488. doi: 10.1016/j.tics.2003.09.002

Whalen, J., Gallistel, C., and Gelman, R. (1999). Non-verbal counting in humans: the psychophysics of number representation. *Psychol. Sci.* 10, 130–137. doi: 10.1111/1467-9280.00120

My brain knows numbers! – an ERP study of preschoolers' numerical knowledge

Tamar Ben-Shalom*, Andrea Berger and Avishai Henik

Department of Psychology and Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer Sheva, Israel

This study investigated brain activity in numerical processing at early stages of development. Brain activity of preschoolers was measured while they performed a numerical Stroop task. Participants were asked to decide which of two digits was numerically or physically larger. Behavioral distance and size congruity effects (SiCEs) were found. However, a reverse facilitation was observed, where responses to neutral trials were faster than to congruent ones. The event-related potentials data showed the expected distance effect at occipitoparietal scalp areas. Moreover, conflict was related to effects both at frontal and parietal scalp areas. In addition, there was a difference between the timing of the interference compared to the facilitation components in the SiCE. In parietal scalp areas, facilitation was significant in an early time window and interference was significant at a later time window. This is consistent with the idea that facilitation and interference are separate processes. Our findings indicate that children as young as 5–6 years old can automatically process the numerical meaning of numerals. In addition, our findings are consistent with the idea that, children might use both frontal and parietal areas in order to process irrelevant numerical information.

Keywords: automatic numerical processing, size congruity effect, ERP, distance effect, brain development

INTRODUCTION

BEHAVIORAL EFFECTS OF NUMERICAL PROCESSING

Our purpose was to investigate how children process numerical values of numerals and how this processing is reflected in their brain activity. Therefore we chose to focus on behavioral effects that indicate automatic processing of numerals. In the study of numerical processing of numerals, two common effects are usually investigated and reported: the *distance effect* (DE) and the *size congruity effect* (SiCE). The DE can be measured when two numerals differ in their numerical value. It was found that subjects are quicker to compare two numerals that are farther apart (e.g., 2 8) than closer ones (e.g., 2 3). This effect was first reported by Moyer and Landauer (1967). The DE was replicated in many studies since then (Henik and Tzelgov, 1982; Dehaene et al., 1990; Tzelgov et al., 1992b). The DE is considered to be an indication for the existence of an analogical mental number line that contains representations of numerals. Representations of close numerals (e.g., 1 2) appear closer on the mental number line than representations of farther apart numerals (e.g., 2 8). Each numeral has its own representative space that overlaps that of the neighboring numerals, and as a result, comparisons are slower for small numerical distances than for large numerical distances.

The SiCE is considered to be evidence for automatic numerical processing. Henik and Tzelgov (1982) were the first to test subjects using the numerical Stroop paradigm. They found that subjects were quicker judging physical sizes when they were congruent to numerical values of the numerals presented (e.g., 3 5). Subjects were slowest when physical sizes and numerical values were incongruent (e.g., 3 5). Henik and Tzelgov considered this as evidence that the numerical dimension is processed in a non-intentional

and automatic manner (Henik and Tzelgov, 1982; Tzelgov et al., 1992b; Rubinsten et al., 2002). Also, by adding a neutral stimulus to the task (e.g., 3 3) it enabled dividing the congruity effect into two components – the *interference* component (incongruent minus neutral trial reaction times, RTs) and *facilitation* component (neutral minus congruent trial RTs). This way, one can examine whether the congruity effect is mostly created by the interference of the irrelevant dimension (the numerical dimension, which the subjects were asked to ignore), or by the facilitation of the irrelevant dimension.

Studies that explored the components of the Stroop effect (i.e., facilitation and interference) tried to dissociate these two components in order to understand what they represent (Posner, 1978; MacLeod and Dunbar, 1988; Tzelgov et al., 1992a; Lindsay and Jacoby, 1994; Szűcs and Soltész, 2007, 2008). Rubinsten and Henik (2006) suggested that the facilitatory component is supposed to involve processes that are more automatic because they are less subjected to strategic control (e.g., see Tzelgov et al., 1992a). Posner (1978) also suggested that facilitation is an indicator of automaticity, whereas interference might reflect attentional processing.

THE DEVELOPMENT OF BEHAVIORAL EFFECTS OF NUMERICAL PROCESSING

Studies that investigated the *development* of the DE and the SiCE and its components (interference and facilitation) found that young children (preschoolers) already showed some of these effects. The DE was found among 5- to 8-year-olds (Sekular and Mierkiewicz, 1977; Duncan and McFarland, 1980; Temple and Posner, 1998; Rubinsten et al., 2002; De Smedt et al., 2009;

Holloway and Ansari, 2009). The results indicated that 5-year-olds already have a mental number line and can use it when necessary. As for the SiCE, this effect was considered to appear only among first or second graders (Girelli et al., 2000; Rubinsten et al., 2002; Mussolin and Noël, 2007, 2008). Rubinsten et al. (2002) studied the development of the components of the SiCE: interference and facilitation. They found that children at the beginning of first grade did not present either of these two components in the physical task (when the numerical dimension was irrelevant). However, children at the end of first grade presented a significant interference effect but not facilitation in the physical task (when the numerical value was irrelevant). This finding might indicate that the facilitation component is more automatic than the interference component and hence appears later among young children, when the automatic numerical processing is more stable and even automatic in its nature. To our knowledge, up until now only one study by Zhou et al. (2007) demonstrated the SiCE at younger ages – among preschoolers (5–6 years old). They related their results to cultural differences between Chinese children and the population of children that have been studied to date.

In a previous behavioral study (Ben Shalom et al., unpublished), we found a significant SiCE among preschoolers (5–6 years old), which showed significant interference and a significant reverse facilitation (see **Figure 1** for similar results). We related this pattern of results to non-mature numerical processing of numerals. These children could automatically relate to the incongruity between physical size and numerical dimensions, and therefore presented an interference effect, but they could not automatically relate to the congruity between the physical and numerical dimensions, and therefore presented a reversed facilitation, meaning that the neutral trials were the easiest for them to judge. After receiving these novel results, we were interested in examining the brain mechanisms behind the numerical processing of these young children.

BRAIN ACTIVITY OF NUMERICAL PROCESSING AND ITS DEVELOPMENT
Neuroimaging studies found that specific parietal areas are activated during numerical processing tasks, and more specifically,
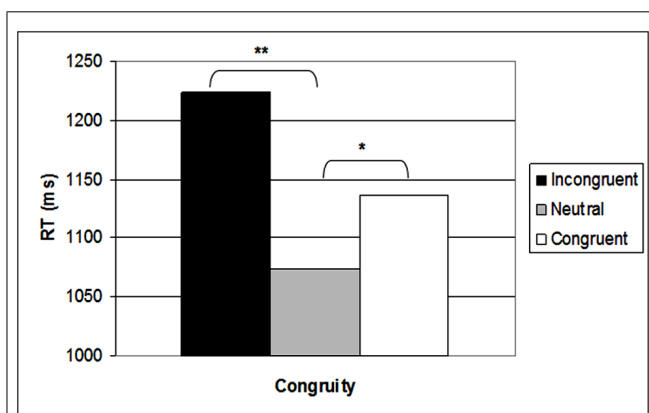
FIGURE 1 | Behavioral SiCE in the physical task. *$p < 0.05$ (two-tailed), **$p < 0.01$ (two-tailed).

the intraparietal sulcus (IPS; Dehaene et al., 2003, 2004; Pinel et al., 2004). This area was found to be modulated by the numerical difference between numerals (i.e., DE; Dehaene et al., 2003; Pinel et al., 2004) and by the incongruity between numerical and physical dimensions (i.e., SiCE; Cohen Kadosh et al., 2007).

Several studies have examined the development of these brain mechanisms of numerical processing. Using functional magnetic resonance imaging (fMRI), Ansari et al. (2005) examined developmental differences in functional neuroanatomy of symbolic number processing. Their results indicated that children's numerical DE was found significant in frontal areas, whereas adults showed this effect in parietal areas. They concluded that this might be an indication for ontogenetic shift throughout development, toward greater parietal engagement in symbolic numerical comparison. Kaufmann et al. (2005) investigated the numerical Stroop effect in children compared to adults. They found that in adult brains, the congruity effect in the numerical Stroop task was seen in the dorsolateral prefrontal cortex and anterior cingulated cortex, and was related to attentional control. Additionally, a larger distance between numerosities resulted in a greater activation in bilateral parietal areas, including the IPS. Kaufmann et al. (2006) also found that the same task activated different brain areas in 9-year-old children. Brain areas that were activated when there was a large numerical distance were frontal but not parietal for the children group. Also, when the numerical value was irrelevant (in the physical task), frontal areas were more activated when comparing the incongruent stimulus activation to the neutral one.

Event-related potential (ERP) studies also investigated the course of the development of the DE and SiCE. Dehaene (1996) found voltage differences associated with numerical distance in a comparison of digits task among adult subjects. This effect was found in the time window of 174–230 ms after stimulus presentation, in electrodes of the occipito–parieto junction. Temple and Posner (1998) examined the development of this brain activity in relation to numerical distance in children. Their study replicated Dehaene's results regarding adults' ERP topography, although they found the effect in an earlier time window (124–234 ms after stimulus presentation). They also revealed the same voltage differences for numerical distance were found among 5-year-old children in a numeral comparison task, although they were slightly delayed compared to adults (around 50 ms after the adults' window). Although it should be mentioned, that in this study, the small and large numerical distance conditions were not perceptually balanced. Hence, ERPs may have been affected by perceptual effects.

Szücs et al. (2007) also examined the development of brain activity during numerical processing. They examine adults and 9- to 11-year-old children using the numerical Stroop task. They replicated results of previous studies finding that both children and adults demonstrated significant voltage differences for numerical distance, between 140 and 320 ms after stimulus presentation and mostly over right parietal electrodes. These findings suggested that children and adults can access the representation of the number line at a similar speed. Looking into brain activity of interference and facilitation components in physical comparisons, Szücs et al. (2007) also found different brain activity patterns for children

compared to adults. Specifically, two wave components – the P300 and lateralized readiness potential (LRP) – were found to be different between those age groups. The interference effect of the Stroop task was more related to the LRP wave (i.e., response conflict) component for children compared to adults. In their opinion, this result indicated that children's slower response in a numerical comparison task was due to an unorganized behavioral response. They concluded that different cognitive processes underlie children's performance in the numerical Stroop task due to non-matured executive function that is required to carry out this task. Soltész et al. (2011b) examined the SiCE in physical judgment among children in grades 1–3. In their article, they stated that they expected stronger interference effects in younger children due to the relatively immature functioning in the prefrontal cortex, but that interference might weaken in favor of facilitation in older children as number processing became more and more automatic. This also supports our hypothesis about the pattern of interference vs. facilitation that we expected to find among 5- to 6-year-old children, as we found in a previous behavioral study.

THE PRESENT STUDY

The purpose of our study was to examine the development of brain mechanisms of DE and SiCE in children of younger ages than were studied so far, that is, among 5- to 6-year-old children, by using an ERP method. According to previous findings, we hypothesized the following:

- Preschoolers would show automatic numerical processing by expressing a significant behavioral DE in the numerical task and SiCE in the physical task.
- Preschoolers would show significant voltage differences to numerical distance according to previous studies (Dehaene, 1996; Temple and Posner, 1998; Szűcs et al., 2007), in the same time window (124–234 ms after stimulus presentation), in electrodes of the occipito–parieto junction.
- If kindergartners showed a significant behavioral SiCE, the ERP topography of this effect would be more frontal than parietal, according to previous studies (Ansari et al., 2005; Kaufmann et al., 2006; Szűcs et al., 2007; Szűcs and Soltész, 2008; Soltész et al., 2011a). Regarding time window, we did not have a specific time window from previous literature because our study is the first one that we are familiar with, that investigated ERP patterns of the SiCE in such young ages.

METHOD
Participants

Seventeen preschoolers – eight males and nine females – aged 5–6 years old (average of 5.5 years old) without any learning or developmental disabilities (based on parental reports) were examined. Children's parents were given payment for their participation. Adult students were given course credit. Parental consent was obtained for the children.

Procedure

Families were contacted through their children's kindergarten. For those families who agreed to participate, a home visit was scheduled to assess the child's IQ and basic numerical abilities. We verified in a home visit that each child knew how to count up to 10 and recognized the numerals 1–9. We also administered the colored RAVEN IQ test in order to measure the child's IQ level. Subsequently, a lab visit was scheduled in which the child performed the numerical Stroop task while his/her brain activity was measured.

The average time for this meeting was 1 h. The task took 20 min (on average) for each child. The order of the tasks was counterbalanced: half of the subjects were tested first on the numerical judgment task and the other half were tested first on the physical judgment task. At the beginning of each task, the participant preformed 12 practice trials where positive feedback for correct answers was given. The experimental blocks themselves did not include any feedback.

At the end of each meeting, parents were given payment for their child's participation. The research was approved by the Israeli Ministry of Education and by the Helsinki Committee.

Stimuli

Two digits appeared in each trial at the center of a computer screen. The distance between the two digits was 10 mm. A typical trial started with a fixation point presented for 300 ms, followed by the two digits that remained in view until the participant pressed one of the computer buttons to indicate which digit was larger. In two separate blocks participants were asked to compare the numerical values of the two digits or their physical sizes. In the numerical block, the two digits differed only in their numerical size and not in their physical size. This was done in order to reduce the amount of trials required for the children's experiment. In the physical block the two digits differed in their numerical and physical sizes. There were only two possible physical sizes: the size of the larger digit was 13 mm and the smaller one was 10 mm.

The stimuli in each block were created using the digits 1, 2, 3, 4, 6, 7, 8, and 9. From these numerals we created two numerical distances: 1 (the pairs: 1–2, 3–4, 6–7, 8–9) and 5 (the pairs: 1–6, 2–7, 3–8, 4–9). Thus, there were eight different stimuli pairs. The pairs could appear with the larger number on the right or with the larger number on the left, allowing for 16 pairs. In the numerical block these stimuli were repeated four times for a total of 64 stimuli. In the physical block, the 16 pairs of stimuli could have each digit appear in two different physical sizes, allowing for 32 different stimuli. The congruent stimuli (e.g., ₃ 8) were repeated as necessary to create 32 congruent stimuli; the same process was carried out to create 32 incongruent stimuli (e.g., 3 ₈). The 32 neutral stimuli were created using a pair of two digits that differed in the physical dimension but not in the numerical dimension (e.g., ₂ 2). Thus, the physical block contained 96 different stimuli.

EEG recording and analysis. E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) was used for stimuli presentation and behavioral data collection. An electroencephalogram (EEG) was recorded from 128 scalp sites using the Electrical Geodesics, Inc. (EGI) Geodesic Sensor Net and system (Eugene, OR; Tucker, 1993). Electrode impedances were kept below 40 kΩ, an acceptable level for this system (Ferree et al., 2001). Data was processed

using a 0.1–100 Hz bandpass filter. Signals were collected at 250 samples per second and digitized with a 24-bit A/D converter. EEG data from trials that were included in the behavioral analysis were processed in Netstation, v4.3 (Electrical Geodesics, Eugene, OR).

Using the Netstation program, continuous EEG data were filtered with a 0.3 Hz high pass and 47 Hz low pass (following Szűcs et al., 2007). The data was then segmented into trials time-locked to the presentation of the stimulus. The length of the segmentation included 100 ms before stimulus presentation and 1,200 ms afterward. Resulting segments were subjected to an automatic bad-channel-, eye blink-, or movement-detection procedure, followed by manual verification. This procedure marks channels as bad if they have a max–min difference higher than 200 μV. It also marks segments with a difference higher than 140 or 55 μV as containing an eye blink or an eye movement, respectively. Segments containing 10 or more bad channels, or those in which any eye activity was detected, were discarded. The minimum number of trials remaining per condition was 25. Before averaging, each trial was re-referenced with the PARE (polar-corrected average reference) re-reference technique for all of the sensors at each time point. Finally, after averaging the trials, subsets were baseline-corrected to 100-ms pre-stimulus presentation and averaged into a grand average of all subjects. Analysis was guided by previous findings of Dehaene (1996); Temple and Posner (1998), and Szűcs et al. (2007), as well as by preliminary visual inspection of the grand-averaged data, using for each effect the difference wave between the conditions. After statistical extraction of the average means for each subject for each condition and time-window, this data was analyzed using a repeated-measures analysis of variance (ANOVA).

Data were analyzed by using repeated-measures ANOVAs for each time window, with numerical distance (1 and 5) and congruity (incongruent, neutral, and congruent) as the within subjects variable. The mean amplitude for the channel group was extracted for each time window for each child in each condition.

The scalp ERP topography that was found for the DE is not fully comparable with the classic location of parietal activity seen in numerical tasks in previous studies (Dehaene, 1996; Temple and Posner, 1998; Szűcs et al., 2007). Therefore, we used preliminary inspection of the difference voltage between the conditions in order to fully capture ERP topography and the time window of the effect. Finally, we used the time window of 284–380 ms after stimulus presentation for the DE, at the ERP topography of the bilateral occipito-parietal area, placing a group of 10 electrodes between P3, P4, O1, and O2 of the 10–20 system. As for the SiCE, visual inspection of the results revealed that children showed voltage differences to congruity conditions in frontal and parietal areas. We again used preliminary inspection of the difference voltage between the conditions. We used three time windows and two ERP topographies: (a) 370–440 ms after stimulus presentation at medial frontal area, placing a group of five electrodes between FZ and CZ of the 10–20 system and (b) 600–750 ms and 810–1190 ms after stimulus presentation at right parietal area, placing a group of five electrodes around P4 of the 10–20 system electrodes.

RESULTS

BEHAVIORAL DATA

Mean RTs were calculated for correct responses. RTs were analyzed as the depended variable. We will divide our results into two sections: DE and SiCE in each task. In the numerical task, the numerical distance was the between subjects variable. In the physical block, the congruity effect was the between subjects variable.

DE in the numerical task

A significant main effect was found for the numerical distance between numerals [$F(1,16) = 7.42$, MSE $= 226{,}780$, $p < 0.01$]. RTs for numerical distance 1 (2,084 ms) were slower than RTs for numerical distance 5 (1,920 ms).

SiCE in the physical task

A significant main effect of congruity was found [$F(2,32) = 17.83$, MSE $= 97{,}019$, $p < 0.001$]. Planned comparisons showed that incongruent trials were significantly slower than congruent trials [$F(1,16) = 19.9$, MSE $= 3{,}253$, $p < 0.001$], and congruent trials were significantly slower than neutral trials [$F(1,16) = 5.82$, MSE $= 5{,}822$, $p < 0.05$]. This created a SiCE with a reverse facilitation (see **Figure 1**; incongruent RT $= 1{,}224$ ms, congruent RT $= 1{,}136$ ms, neutral RT $= 1{,}073$ ms). No significant effect was found for numerical distance.

EEG DATA

Similar to behavioral results, we will divide our results into two sections: DE and SiCE. We found a significant DE, and interference and facilitation effects. As can be seen in **Table 1**, the facilitation appeared to be significant earlier than the interference effect was.

DE in the numerical task

Children showed voltage differences for numerical distance in occipito-parietal areas. The time window that was found significant in the analysis was 284–380 ms after stimulus presentation (see **Figure 2**).

The mean amplitude for the channel group was extracted for each time window for each child in each condition. Data were analyzed by using repeated-measures ANOVAs for each time window, with numerical distance (1 and 5) as the within subjects variable. The ANOVA analysis revealed a significant effect of numerical distance [$F(1,16) = 4.38$, MSE $= 6.2$, $p = 0.05$] in the time window of 284–380 ms (see **Figure 2**). Other time windows were analyzed and found to be non-significant.

SiCE in the physical task

Two ERP topographies were defined for the analysis of the SiCE, according to the previous finding of Szűcs et al. (2007) and preliminary inspection of the waveforms: (1) medial–frontal and (2) right parietal. No significant effect was found in the left parietal area. One time window was defined for the medial–frontal area and two time windows were defined for the right parietal area (see **Figures 3** and **4**).

DISCUSSION

Our study examined brain mechanisms of numerical processing among preschoolers. The main results of the study were: (1) a

Table 1 | SiCE effect in frontal and parietal areas.

Localization	Time window (ms)	Effect	F	MSE	p-Value
Medial–frontal; five electrodes between FZ and CZ	370–440	Congruity	(2,32) = 3.23	2.96	=0.053
		Facilitation	NS		
		Interference	(1,16) = 3.23	5.77	<0.05
Right parietal area; five electrodes around P4	600–750	Congruity	(2,32) = 5.95	7.75	<0.01
		Facilitation	(1,16) = 8.01	11.28	<0.01
		Interference	NS		
	810–1,190	Congruity	(2,32) = 5.95	6.48	<0.01
		Facilitation	NS		
		Interference	(1,16) = 9.86	6.11	<0.01

significant behavioral DE and SiCE were found among these children; (2) a significant reverse facilitation was found when the numerical dimension was irrelevant; (3) brain ERP patterns were modulated by numerical distance in occipito-parietal brain areas; and (4) a congruity effect was found in these children in electrodes above frontal and parietal brain areas. We will discuss each result separately.

BEHAVIORAL DE AND SiCE
Children showed differences in RT according to numerical distance in the numerical task. This result replicates previous findings that showed that children as young as 5 years old can access the mental representation of the number line in a direct comparison task between numerals (Sekular and Mierkiewicz, 1977; Duncan and McFarland, 1980; Temple and Posner, 1998; Rubinsten et al., 2002; De Smedt et al., 2009; Holloway and Ansari, 2009).

In addition, we found a significant behavioral SiCE in the physical task. This is only the second time to our knowledge (except for Zhou et al., 2007) that a SiCE was demonstrated among preschoolers. This finding contradicts several previous studies that found a SiCE only in children at later school ages (Girelli et al., 2000; Rubinsten et al., 2002; Mussolin and Noël, 2007, 2008). A possible explanation could be that today, preschool children are more exposed to numerical stimuli than they were in the past. At least in our country, the current preschool curriculum includes learning of numerals and their association to numerical magnitude. Our behavioral findings indicate that this numerical processing already reached some level of automaticity at this young age.

However, the pattern of the SiCE that the children showed in their RT is still not a fully mature one that characterizes older children and adults in the numerical Stroop task. At preschool age, the children show an inverse facilitation. This means that the RTs to the neutral trials in the physical comparison (e.g., 3 3) were faster than to congruent and incongruent trials. We have already found and reported this pattern in a separate larger sample (Ben Shalom et al., unpublished). Previous findings, such as those of Rubinsten et al. (2002), found that children at the beginning of first grade did not show any facilitation or interference in the physical comparison task. At the end of first grade, children presented an interference component with no facilitation. In our study, the

SiCE was significant in the physical task. This suggests that children at this age already have automatic processing of numerical values.

Interestingly, the pattern of the SiCE that we found indicates a reverse facilitation in the physical judgment. RTs to the neutral trials in the physical comparison (e.g., 3 3) were faster than RTs to the congruent and incongruent trials. According to this pattern, the neutral trials were easier to respond to than the congruent and incongruent trials were. One possible explanation for this pattern is the idea that an additional conflict is involved in the numerical Stroop task. Goldfarb and Henik (2007) suggested that Stroop stimuli create two kinds of conflicts – a task conflict and an information conflict. The task conflict is created because there are two tasks that can be applied to the stimulus – naming the color and reading the (irrelevant) word. The information conflict is created because the stimulus carries information along two dimensions – the information provided by the meaning of the word and the information provided by the color of the word. The incongruent pairs create both types of conflicts (information and task). In contrast, the congruent pairs present only the task conflict because the information from both dimensions point in the same direction. The neutral pairs have no conflict at all.

Goldfarb and Henik (2007) reduced cognitive control, and found a reverse facilitation – the neutral trials were faster than congruent trials. Their explanation was that this happened because they revealed the task conflict in the congruent trials. In light of this study it is interesting to find a reverse facilitation in young children and in numerical cognition. The reverse facilitation in our study was not due to manipulation of control. We suggest that this reverse facilitation in kindergarten children is due to premature control ability. Specifically, the children had to switch from comparing the physical sizes to comparing the numerical values (or vice versa). The cognitive ability to switch and manage conflicts is probably not fully developed in these young children. Hence, the congruent condition is still more difficult for them because it contains a conflict (the task conflict). Also, Children in this age group are well trained in differentiating physical sizes, so a comparison between sizes becomes automatic, even when irrelevant, and can be processed fast enough to interfere with or facilitate numerical comparisons. On the other hand, our
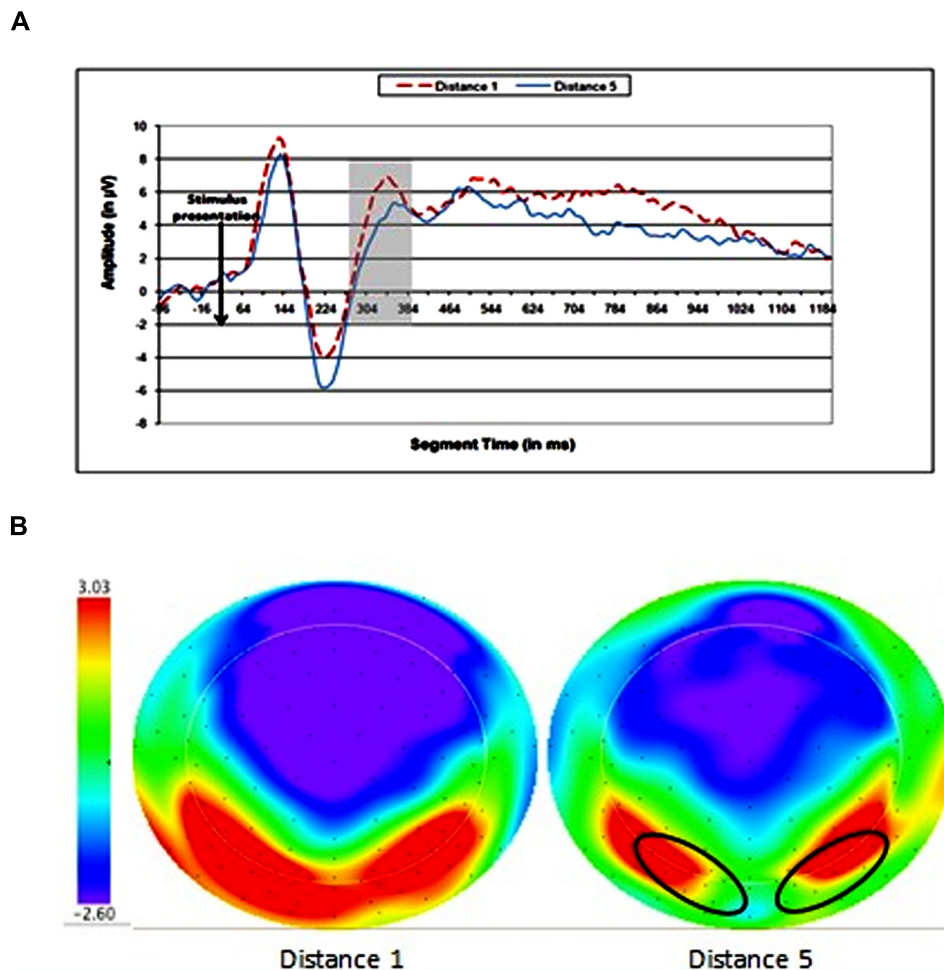
FIGURE 2 | Electrophysiological results for the numerical distance effect. (A) Wave forms of distances 1 and 5 at occipitoparietal electrode sites. **(B)** Topographic maps of scalp voltages at around 330 ms after stimulus presentation, for distances 1 and 5. A black circle indicates the area of electrodes that were analyzed.

group of children was not yet well-trained, and thus, they did not automatically respond to the numerical values of numerals. This dimension can create conflict and interference, but it does not facilitate a child's decision in congruent comparisons. This explanation was also given in Rubinsten et al. (2002). They investigated (among other age groups) first graders in the numerical Stroop task. They found that in this group, in the physical task (when the numerical dimension was irrelevant), size congruity was composed only of the interference component at the beginning of first grade. The facilitation component appeared later, in the older age groups. They concluded that the interference component, which is more automatic in nature, appeared earlier than the facilitation component in the physical task (when numerical values were irrelevant).

Another possible explanation for the reverse facilitation pattern observed in our study relies on the difference in the capability of the children at this age to attend physical sizes, as opposed to numerical sizes. The processing of numerical value, in contrast to physical size, is not
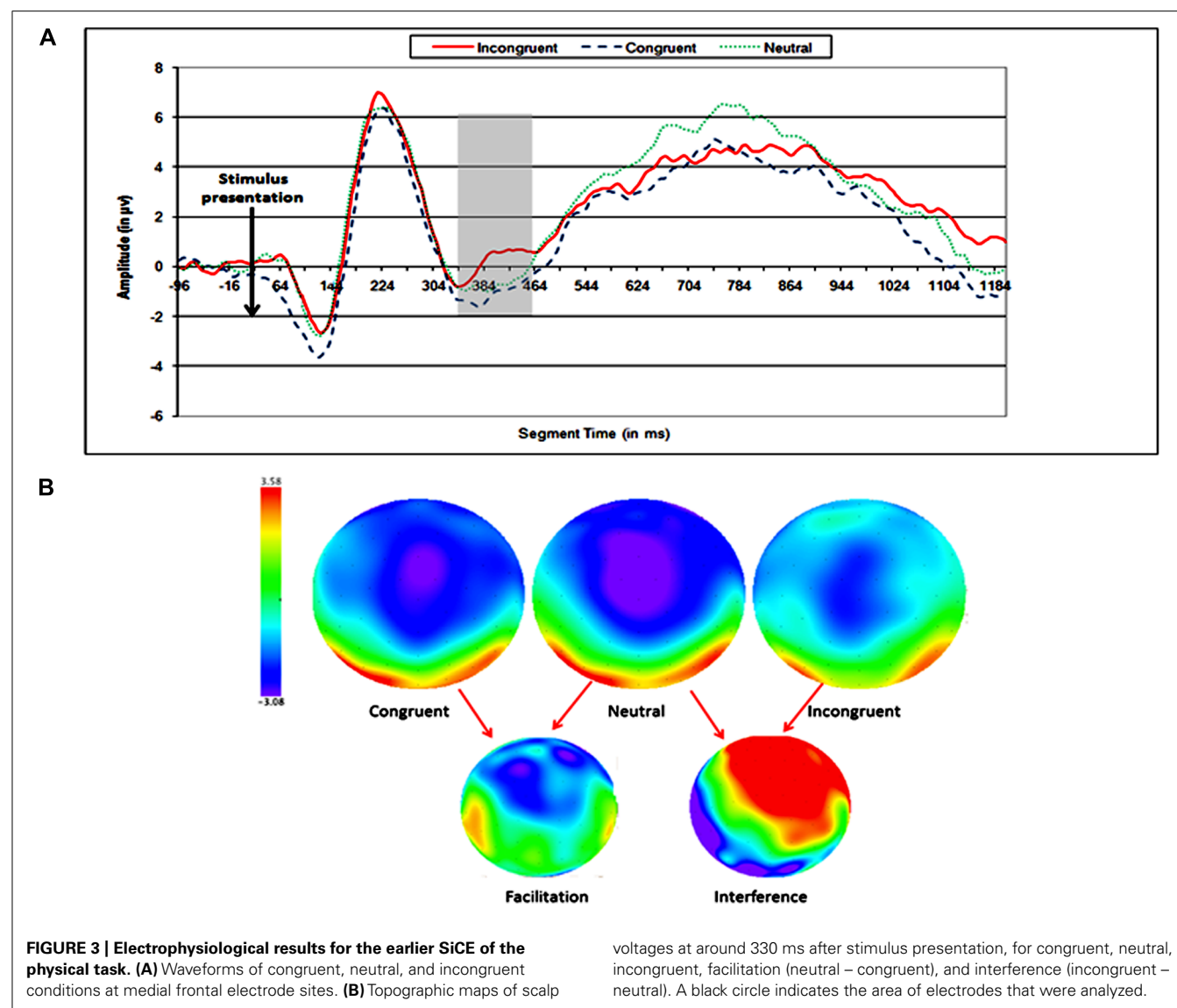
enough trained and automatic. Thus, when it is irrelevant, although it creates conflict and interferes, it is not processed fast enough to facilitate a child's response in congruent trials.

In either case, the reverse facilitation pattern seems to reflect a still relatively immature processing of the numerical dimension of the stimuli, when this dimension is irrelevant to the task. Interestingly, a similar pattern of lack of facilitation or "reverse facilitation" has also been reported (Rubinsten et al., 2002; Rubinsten and Henik, 2006; Ashkenazi et al., 2008).

OCCIPITO-PARIETAL EFFECTS OF NUMERICAL DISTANCE
We found a significant DE in data gathered from electrodes above the occipito-parietal area in the time window 284–380 ms after stimulus presentation. Our time window for the DE is similar to that reported by Szűcs et al. (2007; i.e., 240–320 ms). Moreover, our topography of the DE (occipito-parietal junction) is very similar to Szűcs et al.'s 2007 ERP topography in grade 3 children (the youngest subjects in their study). Our

FIGURE 3 | Electrophysiological results for the earlier SiCE of the physical task. (A) Waveforms of congruent, neutral, and incongruent conditions at medial frontal electrode sites. **(B)** Topographic maps of scalp voltages at around 330 ms after stimulus presentation, for congruent, neutral, incongruent, facilitation (neutral – congruent), and interference (incongruent – neutral). A black circle indicates the area of electrodes that were analyzed.

results partially replicate those of Temple and Posner (1998). However, they found a bi-lateral DE and our time window for this effect is somewhat later. These differences probably result from the fact that in their study the children performed a straightforward numerical comparison, while in our study the children were presented with two competing dimensions.
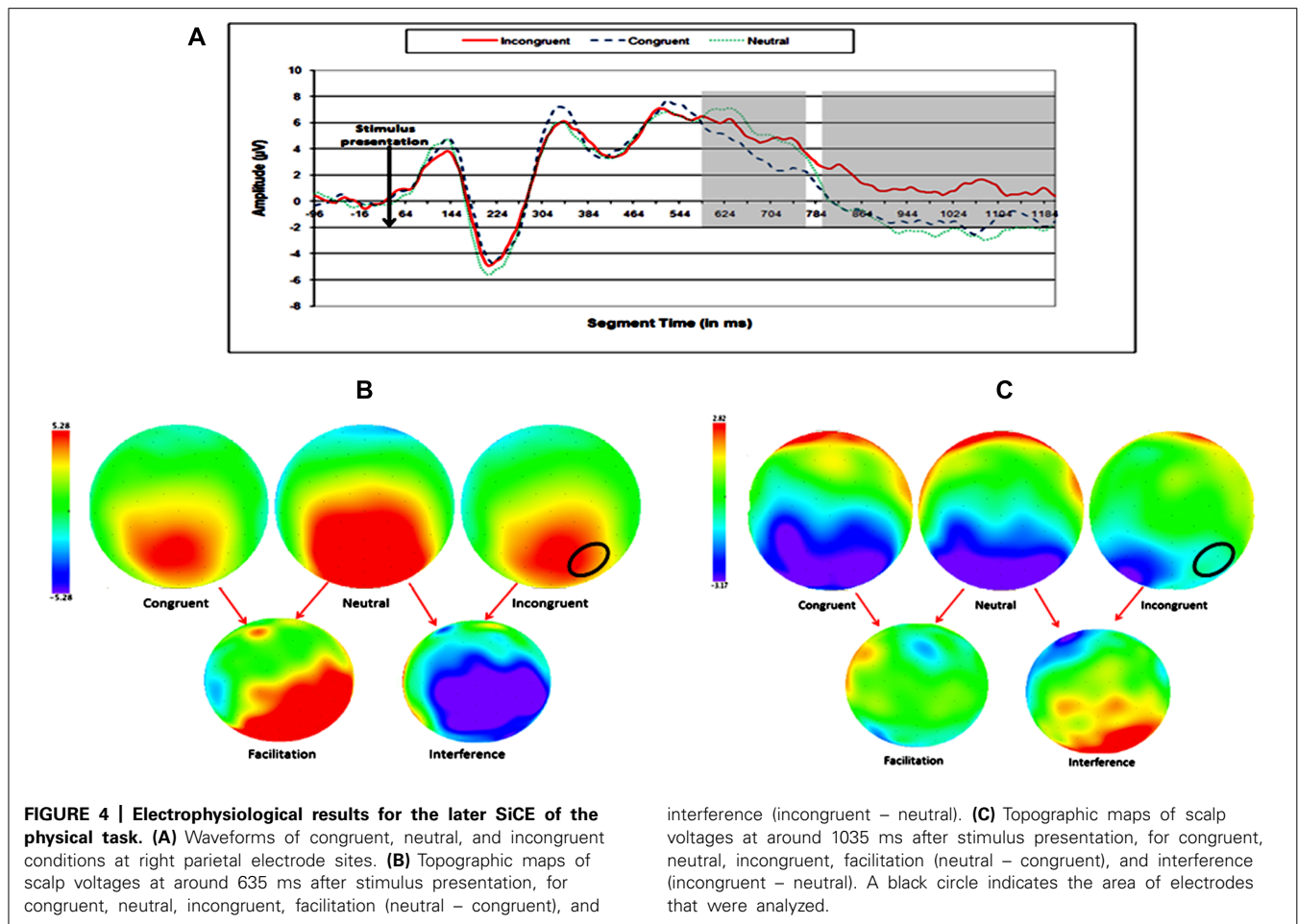
FRONTAL AND PARIETAL EFFECTS IN THE SiCE

In our study, the SiCE effect was found in data gathered from electrodes above both frontal and parietal areas. Interestingly, the timing of the congruity effect was earlier in frontal electrodes and later in right parietal electrodes. This result could indicate an earlier detection of the incongruity between the physical and numerical dimension by the frontal lobe. The significant later effect among parietal electrodes could indicate a more detailed processing of the conflict presented by the irrelevant dimension, and the automatic activation of the numerical dimension. Studies

have found that children showed more frontal ERP effect when asked to perform numerical comparison (Ansari et al., 2005; Kaufmann et al., 2005, 2006). This frontal ERP effect was related to immature numerical processing that is based more on the executive function system. Our results support this idea, as seen in the early activation of the executive function system (in more frontal areas). However, we found later ERP effect above the right parietal area, which is considered in many studies as the area that is responsible for numerical processing, especially in the numerical Stroop task (Dehaene et al., 2003; Pinel et al., 2004; Cohen Kadosh et al., 2007). Our findings are novel in light of the fact that no research to date has found parietal ERP effect in children at such a young age using the numerical Stroop task.

Another aspect of our results is that they lend support to the research of Szücs and Soltész (2007, 2008) by differentiating the facilitation and interference components of the SiCE. In our study, as well as in theirs, the facilitation over electrodes above parietal

FIGURE 4 | Electrophysiological results for the later SiCE of the physical task. (A) Waveforms of congruent, neutral, and incongruent conditions at right parietal electrode sites. **(B)** Topographic maps of scalp voltages at around 635 ms after stimulus presentation, for congruent, neutral, incongruent, facilitation (neutral – congruent), and interference (incongruent – neutral). **(C)** Topographic maps of scalp voltages at around 1035 ms after stimulus presentation, for congruent, neutral, incongruent, facilitation (neutral – congruent), and interference (incongruent – neutral). A black circle indicates the area of electrodes that were analyzed.

areas appeared earlier than the interference did. In Szűcs and Soltész's study they relate the facilitation component to the processing stage of the stimuli (regarding the numerical and physical sizes) and the interference to response selection. Our results cannot clearly suggest the same idea, but the time course of parietal activation is similar.

CONCLUSION

During the last year of kindergarten, children already show an automatic activation of the numerical value of numerals. The pattern of RTs at this age is unique in the physical task (when the numerical value is irrelevant; Ben Shalom et al., unpublished). This pattern of results resembles the pattern of the SiCE of discalculic adults or acalculic patients (Rubinsten et al., 2002; Rubinsten and Henik, 2006; Ashkenazi et al., 2008). In addition, young children show brain activation sensitivity, in ERP effects over frontal and parietal areas, to the numerical distance between numerals and to incongruity between the numerical and physical dimensions. The early activation found in the data gathered from electrodes above frontal areas can be related to conflict management that these children needed to activate in order to process the incongruity between dimensions. The later activation among electrodes above parietal areas could indicate that even more mature networks in the parietal area were activated

later and used to process the numerical information. This is the first study to our knowledge that showed this kind of brain activation of the SiCE in children at such young ages (6- to 5-year-olds). Other studies that previously examined the automatic activation of the numerical dimension found this effect only in older-aged children (first–second grade). However, a speculative hypothesis can be made about cohort differences between children who are studied today as opposed to children who were studied in the past. Today in Israel, preschoolers learn the numerals 1–10 and the association between numerals and quantities in a formal way. This can explain the difference between our results and previous results in the literature. However, more research needs to be done in order to expand our and Zhou et al.'s2007 results regarding preschoolers' numerical automatic processing in order to fully understand this effect and its relation to individual differences in children's numerical abilities.

ACKNOWLEDGMENTS

REFERENCES

Ansari, D., Garcia, N., Lucas, E., Hamon, K., and Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *Neuroreport* 16, 1769–1773. doi: 10.1097/01.wnr.0000183905.23396.f1

Ashkenazi, S., Henik, A., Ifergane, G., and Shelef, I. (2008). Basic numerical processing in left intraparietal sulcus (IPS) acalculia. *Cortex* 44, 439–448. doi: 10.1016/j.cortex.2007.08.008

Cohen Kadosh, R., Cohen Kadosh, K., Linden, D. E. J., Gevers, W., Berger, A., and Henik, A. (2007). The brain locus of interaction between number and size: a combined functional magnetic resonance imaging and event-related potential study. *J. Cogn. Neurosci.* 19, 957–970. doi: 10.1162/jocn.2007.19.6.957

Dehaene, S. (1996). The organization of brain activations in number comparison: an event-related potentials and the additive-factors method. *J. Cogn. Neurosci.* 8, 47–86. doi: 10.1162/jocn.1996.8.1.47

Dehaene, S., Dupoux, E., and Mehler, J. (1990). Is numerical comparison digital: analogical and symbolic effect in two-digit number comparison. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 626–641. doi: 10.1037/0096-1523.16.3.626

Dehaene, S., Molko, N., Cohen, L., and Wilson, A. (2004). Arithmetic and the brain. *Curr. Opin. Neurobiol.* 14, 218–224. doi: 10.1016/j.conb.2004.03.008

Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239

De Smedt, B., Verschaffel, L., and Pol Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *J. Exp. Child Psychol.* 103, 469–479. doi: 10.1016/j.jecp.2009.01.010

Duncan, E. M., and McFarland, C. E. (1980). Isolating the effect of symbolic distance and semantic congruity in comparative judgments: an additive-factors analysis. *Mem. Cogn.* 2, 95–110.

Ferree, T. C., Luu, P., Russell, G. S., and Tucker, D. M. (2001). Scalp electrode impedance, infection risk, and EEG data quality. *Clin. Neurophysiol.* 112, 536–544. doi: 10.1016/S1388-2457(00)00533-2

Girelli, L., Lucangeli, D., and Butterworth, B. (2000). The development of automaticity in accessing number magnitude. *J. Exp. Child Psychol.* 76, 104–122. doi: 10.1006/jecp.2000.2564

Goldfarb, L., and Henik, A. (2007). Evidence for task conflict in the Stroop effect. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1170–1176. doi: 10.1037/0096-1523.33.5.1170

Henik, A., and Tzelgov, J. (1982). Is three greater than five: the relation between physical and semantic size in comparison tasks. *Mem. Cogn.* 10, 389–395. doi: 10.3758/BF03202431

Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children's mathematics achievement. *J. Exp. Child Psychol.* 103, 17–29. doi: 10.1016/j.jecp.2008.04.001

Kaufmann, L., Koppelstaetter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., et al. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: an event-related fMRI study. *Neuroimage* 25, 888–898. doi: 10.1016/j.neuroimage.2004.12.041

Kaufmann, L., Koppelstaetter, F., Siedentopf, C., Haala, I., Haberlandt, E., Zimmerhackl, L. B., et al. (2006). Neural correlates of the number-size interference task in children. *Neuroreport* 17, 587–591. doi: 10.1097/00001756-200604240-00007

Lindsay, D. S., and Jacoby, L. L. (1994). Stroop process dissociation: the relationship between facilitation and interference. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 219–234. doi: 10.1037/0096-1523.20.2.219

MacLeod, C. M., and Dunbar, K. (1988). Training and Stroop-like interference: evidence of continuum of automaticity. *J. Exp. Psychol. Learn. Mem. Cogn.* 14, 126–135. doi: 10.1037/0278-7393.14.1.126

Moyer, R. S., and Landauer, T. K. (1967). Time required for judgment of numerical inequality. *Nature* 215, 1519–1520. doi: 10.1038/2151519a0

Mussolin, C., and Noël, M. P. (2007). The nonintentional processing of Arabic numbers in children. *J. Clin. Exp. Neuropsychol.* 29, 225–234. doi: 10.1080/13803390600629759

Mussolin, C., and Noël, M. P. (2008). Automaticity for numerical magnitude of two-digit Arabic numbers in children. *Acta Psychol.* 129, 264–272. doi: 10.1016/j.actpsy.2008.08.001

Pinel, P., Piazza, M., and Bihan, D. L. (2004). Distributed and overlapping cerebral representation of number, size, and luminance during comparative judgments. *Neuron* 41, 983–993. doi: 10.1016/S0896-6273(04)00107-2

Posner, M. I. (1978). *Chronometric Explorations of Mind*. Hillsdale, NJ: Erlbaum.

Rubinsten, O., and Henik, A. (2006). Double dissociation of functions in developmental dyslexia and dyscalculia. *J. Educ. Psychol.* 98, 854–967. doi: 10.1037/0022-0663.98.4.854

Rubinsten, O., Henik, A., Berger, A., and Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *J. Exp. Child Psychol.* 81, 74–92. doi: 10.1006/jecp.2001.2645

Sekular, R., and Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Dev.* 48, 630–633. doi: 10.2307/1128664

Soltész, F., Goswami, U., White, S., and Szücs, D. (2011a). Executive function effects and numerical development in children: behavioural and ERP evidence from a numerical Stroop paradigm. *Learn. Individ. Differ.* 21, 662–671. doi: 10.1016/j.lindif.2010.10.004

Soltész, F., White, S., and Szücs, D. (2011b). Event-related brain potentials dissociate the developmental time-course of automatic numerical magnitude analysis and cognitive control functions during the first three years of primary school. *Dev. Neuropsychol.* 36, 682–701. doi: 10.1080/87565641.2010.549982

Szücs, D., and Soltész, F. (2007). Event related potentials dissociate facilitation and interference effects in the numerical Stroop paradigm. *Neuropsychologia* 45, 3190–3202. doi: 10.1016/j.neuropsychologia.2007.06.013

Szücs, D., and Soltész, F. (2008). The interaction of task-relevant and task-irrelevant stimulus features in the number/size congruency paradigm: an ERP study. *Brain Res.* 1190, 143–158. doi: 10.1016/j.brainres.2007.11.010

Szücs, D., Soltész, F., Jármi, E., and Csépe, V. (2007). The speed of magnitude processing and executive functions in controlled and automatic number comparison in children: an electro-encephalography study. *Behav. Brain Funct.* 3, 23. doi: 10.1186/1744-9081-3-23

Temple, E., and Posner, M. I. (1998). Brain mechanisms of quantity are similar in 5-year-olds and adults. *Proc. Natl. Acad. Sci. U.S.A.* 95, 7836–7841. doi: 10.1073/pnas.95.13.7836

Tucker, D. M. (1993). Spatial sampling of head electrical fields: the geodesic sensor net. *Electroencephalogr. Clin. Neurophysiol.* 87, 154–163. doi: 10.1016/0013-4694(93)90121-B

Tzelgov, J., Henik, A., and Berger, J. (1992a). Controlling Stroop effects by manipulating expectation for colored words. *Mem. Cogn.* 20, 727–735. doi: 10.3758/BF03202722

Tzelgov, J., Meyer, J., and Henik, A. (1992b). Automatic and intentional processing of numerical information. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 166–179. doi: 10.1037/0278-7393.18.1.166

Zhou, X., Chen, Y., Chen, C., Jiang, T., Zhang, H., and Dong, Q. (2007). Chinese kindergartners' automatic processing of numerical magnitude in Stroop-like tasks. *Mem. Cogn.* 35, 464–470. doi: 10.3758/BF03193286

A componential view of children's difficulties in learning fractions

Florence Gabriel[1,2]*, Frédéric Coché[3], Dénes Szucs[1]*, Vincent Carette[3], Bernard Rey[3] and Alain Content[2]

[1] Department of Experimental Psychology, Centre for Neuroscience in Education, University of Cambridge, UK
[2] Laboratoire Cognition, Langage et Développement, Centre de Recherche Cognition et Neurosciences, Université Libre de Bruxelles (ULB), Bruxelles, Belgium
[3] Service des Sciences de l'Education, Faculté des Sciences Psychologiques et de l'Education, Université Libre de Bruxelles (ULB), Bruxelles, Belgium

Fractions are well known to be difficult to learn. Various hypotheses have been proposed in order to explain those difficulties: fractions can denote different concepts; their understanding requires a conceptual reorganization with regard to natural numbers; and using fractions involves the articulation of conceptual knowledge with complex manipulation of procedures. In order to encompass the major aspects of knowledge about fractions, we propose to distinguish between conceptual and procedural knowledge. We designed a test aimed at assessing the main components of fraction knowledge. The test was carried out by fourth-, fifth- and sixth-graders from the French Community of Belgium. The results showed large differences between categories. Pupils seemed to master the part-whole concept, whereas numbers and operations posed problems. Moreover, pupils seemed to apply procedures they do not fully understand. Our results offer further directions to explain why fractions are amongst the most difficult mathematical topics in primary education. This study offers a number of recommendations on how to teach fractions.

Keywords: fractions, equivalence, part-whole, proportion, arithmetic operations, fraction subcontructs

INTRODUCTION

As the joke goes, "three out of two people have trouble with fractions." Fractions have been known from ancient civilizations until current times, but they still pose major problems when learning mathematics. Babylonian civilization and Egyptians of 4000 years ago already worked with fractions. The processing of fractions is part of our everyday life and is used in situations such as the estimation of rebates, following a recipe or reading a map. Moreover, fractions play a key role in mathematics, since they are involved in probabilistic, proportional and algebraic reasoning. Then why is it so hard for pupils to learn and represent fractions? Fractions have been used for centuries and are manipulated in a great variety of everyday life situations and in mathematics, and yet they are hard for students to grasp and master. In this article, we will try to shed light on children's difficulties when they learn fractions.

Fractions are well-known to constitute a stumbling block for primary school children (Behr et al., 1983; Moss and Case, 1999; Grégoire and Meert, 2005; Charalambous and Pitta-Pantazi, 2007). Understanding difficulties in learning fractions seems absolutely crucial as they can lead to mathematics anxiety, and affect opportunities for further engagement in mathematics and science. Various hypotheses have been proposed in order to explain those difficulties. In this research, we used a theoretical framework based on psychological and educational theories to define problems encountered by pupils when they learn fractions. We tested 4th, 5th, and 6th-graders in order to identify children's difficulties more precisely.

DIFFERENT OBSTACLES IN LEARNING FRACTION
Whole number bias

Fractions are rational numbers. A rational number can be defined as a number expressed by the quotient a/b of integers, where the denominator, b, is non-zero. According to a recent theory of numerical development, children who have not yet learned fractions generally believe that the properties of whole numbers are the same for all numbers (Siegler et al., 2011). Indeed, one of the main difficulties when learning fractions comes from the use of natural number properties to make inferences on rational numbers, what Ni and Zhou (2005) called the "whole numbers bias." This bias leads to difficulties conceptualizing whole numbers as decomposable units.

From a mathematical viewpoint, there are fundamental differences between those two types of numbers. Firstly, rational numbers are a densely ordered set, whereas whole numbers form a discrete set. Between two rational numbers, there is an infinity of other rational numbers, while between two natural numbers, there is no other natural number (Vamvakoussi and Vosniadou, 2004). Secondly, another feature of rational numbers is the possibility to write them from an infinity of fractions. This corresponds to the notion of equivalent fractions. Thirdly, faction symbols are a/b types. Pupils often process numerator and denominator as two separate whole numbers (Pitkethly and Hunting, 1996). They apply procedures that can only be used with whole numbers (Nunes and Bryant, 1996). Consequently, typical errors appear in addition or subtraction tasks (e.g., 1/4 + 1/2 = 2/6), and also in

fraction comparison (e.g., 1/5 >1/3). In this case, pupils' reasoning can be resumed as follows: if the number is larger, then the magnitude it represents is larger. But when we think about fractions, a larger denominator does not mean a larger magnitude, but a smaller one. Another difficulty appears in multiplication tasks. Multiplying natural numbers always lead to a larger answer, but it is not the case with fractions (e.g., $8 \times 1/4 = 2$).

The inappropriate generalization of the knowledge about natural numbers is even more resistant as it is widely anterior to the one about rational numbers (Vamvakoussi and Vosniadou, 2004). In order to overcome these mistakes, it would seem necessary for students to perform a conceptual reorganisation which integrates rational numbers as a new category of numbers, with their own rules and functioning (Stafylidou and Vosniadou, 2004). Furthermore, even in adults, knowledge about natural numbers is often preponderant when processing fractions (Bonato et al., 2007; Kallai and Tzelgov, 2009).

Different meanings of fractions

Another major difficulty comes from the multifaceted notion of fractions (Kieren, 1993; Brousseau et al., 2004; Grégoire and Meert, 2005). Kieren (1976) was the first to separate fractions into four interrelated categories: ratio; operator; quotient; and measure. The ratio category expresses the notion of a comparison between two quantities, for example when there are three boys for every four girls in a group. So in this case, the ratio of boys to girls is 3:4; the boys representing 3/7 of the group and the girls 4/7 of the group. In the operator category, fractions are considered as functions applied to objects, numbers or sets (Behr et al., 1983). The fraction operator can enlarge or shrink a quantity to a new value. For example, finding 3/4 of a number can be a function where the operation is multiply by 3 divided by 4, or divided by 4 and then multiply by 3. The quotient category refers to the result of a division. For example, the fraction 3/4 may be considered as a quotient, 3/4. In the measure category, fractions are associated with two interrelated notions. Firstly, they are considered as numbers, which convey how big the fractions are. Secondly, they are associated with the measure of an interval. According to Kieren (1976), the part-whole notion of fractions is implicated in these four categories. That is the reason why he did not describe it as a fifth category.

Thereafter, Behr et al. (1983) proposed a theoretical model linking the different categories of fractions. They recommend considering part-whole as an additional category. They also associated partitioning to the part-whole notion. The part-whole category can then be defined as a situation in which a continuous quantity is partitioned into equal size (e.g., dividing a cake into equal parts), and partitioning would be the same with a set of discrete objects (e.g., distributing the same amount of sweets among a group of children).

Other models have been proposed to describe the multiple meanings of fractions (Brissiaud, 1998; Rouche, 1998; Mamede et al., 2005). These models partly overlap, but are not entirely equivalent. For instance, Mamede et al. (2005) present four types of fraction use: quantifying a part-whole relationship, quantifying a quotient, representing an operator, representing a relation between quantities. Meanwhile Grégoire (2008) suggests a different model, in which three categories correspond to three acquisition stages. In the first stage, the fraction is seen as an operator. This notion refers to sharing situations. The second one is the ratio stage which requires a high level of abstraction because one needs to understand that different fractions can represent the same ratio. This is linked to the notion of equivalent fractions. The third and last stage is related to the numerical meaning of fractions. Fractions are here conceived as a new category of numbers, with their own rules and properties.

Conceptual and procedural understanding

Another explanation of children's difficulties when learning fractions lies in the articulation between conceptual and procedural knowledge. Previous studies have shown that children would often perform calculations without knowing why (Kerslake, 1986).

Conceptual knowledge can be defined as the explicit or implicit understanding of the principles ruling a domain and the interrelations between the different parts of knowledge in a domain (Rittle-Johnson and Alibali, 1999). It can also be considered as the knowledge of central concepts and principles, and their interrelations in a particular domain (Schneider and Stern, 2005). Conceptual knowledge is thought to be mentally stored in a form of relational representations, such as semantic networks (Hiebert, 1986). It is not tied to a specific problem, but can be generalized to a class of problems (Hiebert, 1986; Schneider and Stern, 2010).

Procedural knowledge can be defined as sequences of actions that are useful to solve problems (Rittle-Johnson and Alibali, 1999). Some authors consider procedural knowledge as the knowledge of symbolic representations, algorithms, and rules (Byrnes and Wasik, 1991). Moreover, procedural knowledge would allow people to solve problems in a quick and effective way as it can easily be automatized (Schneider and Stern, 2010). Therefore, it can be used with few cognitive resources (Schneider and Stern, 2010). However, procedural knowledge is not as flexible as conceptual knowledge and is often bound to specific problem types (Baroody, 2003).

Those two types of knowledge may not evolve in independent ways. Many theories on knowledge acquisition suggest that the generation of procedures is based on conceptual understanding (Halford, 1993; Gelman and Williams, 1997). They argue that children use their conceptual understanding to develop their discovery procedures and adapt acquired procedures to new tasks. According to this approach, children's difficulties when learning about fractions could be interpreted as a use of mathematical symbols without access to their meaning. Procedural knowledge may also influence conceptual understanding. Using procedures would lead to a better conceptual understanding. But few studies support this idea. For instance, Byrnes and Wasik (1991) argue that many children learn the right procedures to multiply fractions, but they never seem to understand the underlying principles. Other authors support a third point of view. Both types of knowledge might progress in an iterative and interactive way (Rittle-Johnson et al., 2001). Conceptual and procedural knowledge might continually and incrementally stimulate each other. Neither would necessarily precede the other.

In mathematics education, teachers seem to focus more on procedural than conceptual knowledge. Children usually learn rote procedures in a repetitive way. This leads to a misunderstanding of mathematical symbols (Byrnes and Wasik, 1991). Consequently many computational errors are due to an impoverished conceptual understanding.

OUR THEORETICAL FRAMEWORK

Taking into account the different theoretical models presented and the issues they arise led us to build our own conceptual framework. In this study exploring the difficulties in learning fractions, two main components were considered: a conceptual component and a procedural component.

The conceptual component was divided in four distinct aspects: proportion, number, measure and part-whole/partition. Part-whole/partition refers to how much of an object (e.g., 1/2 pizza) or a collection (e.g., 1/2 of a bag of sweets) is represented by the fraction symbol (Hecht et al., 2003; Kieren, 1988). Typical tasks used to assess that kind of conceptual knowledge involve shading parts of a figure indicated by a fraction, or the opposite exercise consisting of writing the fraction representing the quantity of a figure that is shaded (Hiebert and Lefevre, 1986; Byrnes and Wasik, 1991; Ni, 2001). Proportion represents the comparison between two quantities. We used comparison of different expressions of the same ratio (e.g., 1/2, 2/4, and 3/?) as it is an adequate way to assess the understanding of proportion. The numerical meaning of fraction refers to the fact that fractions represent rational numbers that can be ordered on a number line (Kieren, 1988). Two relevant tasks were used to assess children's understanding of the numerical meaning of fractions: firstly, number lines on which they are asked to place a fraction, and secondly, indicating which of several given fractions represents the largest quantity (Byrnes and Wasik, 1991; Ni, 2000).

Several variables also held our attention regarding the representation of fractions. Discrete and continuous quantities were used. Children might have greater difficulties to link 2/4 to 2 out of 4 for elements of a set than 2/4 of a pie (Ni, 2001). Multiple objects and figures, as well as numerical symbols were introduced to assess the possible interference of certain types of representations (Coquin-Viennot and Camos, 2006). For practical reasons, we did not examine fractions as a measure in this study. This category is closely related to the metric system. The manipulation of fractions as a measure can be made by splitting units of length, area, volume, time, mass, etc. Understanding these measuring situations involves several concepts that are not exclusively related to fractions, such as understanding different unit systems or a good grasp of the decimal position system. Therefore, it is difficult to assess the understanding of this category in isolation from these variables.

Procedural items were those that could be easily solved by applying a procedure that could be implemented without checking for meaning outside that particular procedure. The procedural component involved various operations on fractions, namely the addition and subtraction with or without common denominators, multiplication, and simplification of fractions. Children were given different arithmetical operations to solve as well as simplification exercises.

RESEARCH QUESTIONS

The main aim of this study was to provide empirical data that could explain difficulties encountered by children when they learn fractions. Our first objective was to analyse the mathematics curriculum of the French Community of Belgium, where this study was conducted. Our second objective was to understand the nature of pupils' difficulties through different categories.

We addressed several research questions regarding children's difficulties when learning fraction. First, we wanted to define more precisely the difficulties encountered by primary school children. Second, one of the goals of this study was to clarify the relationship between conceptual and procedural knowledge of fractions. Does conceptual knowledge of fractions influence procedural knowledge? Or is procedural knowledge sufficient to understand fractions? Our hypothesis is that children's difficulties come from a lack of conceptual understanding of fractions. Their errors would come from the application of routine procedures, but they do not understand the various underlying concepts.

Conceptual knowledge of fractions was assessed through tests about the different meanings of fractions (part-whole, proportion, number), and the different representations of fractions (e.g., association between figural, numeral, and verbal representations). Procedural knowledge about fractions was evaluated through operations on fractions and simplification tasks.

METHODS

PARTICIPANTS

The test was administered to eight Grade 4 classes (mean age: 9 years 11 months old), eight Grade 5 (mean age: 11 years 1 month old) classes and eight Grade 6 classes (mean age: 12 years old) from five different schools, representing a total sample of 439 participants (214 girls and 225 boys). The choice of these grades was deliberate, as fraction learning usually starts from Grade 4 in the French Community of Belgium where the study was conducted. Informed consent was obtained from parents and the director of every school, as well as from the 24 teachers involved in this research. Assent from children was obtained at the onset of both testing sessions.

THE SETTING OF THE STUDY

We analyzed 21 mathematics textbooks recognized by the Education Department of the French Community of Belgium. Fraction concepts used in mathematics textbooks in Grade 4–6 were listed. The goal was to analyse the progression of fraction learning proposed by those textbooks. The most striking observation was that there was a great variety of ways to introduce fractions. In most textbooks, the part-whole concept was considered as the starting point, but in some cases, the measure concept was introduced first. Every concept described in our theoretical framework was represented in the textbooks, but the number of exercises concerning each one of them varied greatly.

We also examined the official mathematics program of the French Community of Belgium. The program presents, in a structured way, the basic skills for the first 8 years of compulsory education, and the skills pupils have to master by the end of each stage (Ministère de la Communauté française, 1999). Fractions were divided into two different categories, Numbers

and Quantities. Any requirement at the end of primary school (Grade 6) is briefly reviewed in this section. In the Number category, pupils should be able count, enumerate and classify fractions as well as decimal numbers. They should also be able to calculate, identify and solve operations involving fractions and decimal numbers. In the Quantities category, children are supposed to operate and fractionate different quantities in order to compare them. They should be able to add up and subtract two fractions as well as calculating percentages. The program also mentioned their ability to solve proportionality problems.

The official program offers a list of what pupils should know about fractions in primary school. But what did not appear clearly was a logical progression between all the meanings of fractions. For example, how and when should equivalent fractions be introduced? There was not a clear development for teaching fraction. This situation may be risky as teachers might present fractions as a succession of different independent activities with no real underlying logical progression.

In order to complete the information found in the textbooks, we analyzed pedagogical practices about the way teachers introduce and teach fractions. This investigation revealed the great variety of ways to teach fractions. Our analysis was based on different sources. Firstly, we asked the 24 teachers involved in this study to give us a list of all the activities about fractions conducted in their classrooms. Secondly, teachers gave us a sample of their lessons on fractions as well as pupils notebooks. Thirdly, we made informal observations during the tests.

In Grade 4, pupils learn how to read and represent the value of a fraction. They start placing fractions on a graduated number line. They learn how to simplify fractions (i.e., introduction to equivalent fractions). They learn how to add and subtract of fractions with small and common denominators. In Grade 5, children learn more about fractions as numbers and how they represent quantities. Pupils are trained to convert fractions into decimal numbers and vice versa. They use addition and subtraction of fractions with different denominators. Improper fractions are introduced. In Grade 6, multiplication of fractions is introduced.

Our analysis highlighted the fact that teachers are more inclined to use procedures than what is recommended by the official program. The different conceptual meanings are presented successively without any logical progression. The order in which they are introduced depends on the teacher and on the textbook used by the teacher. Furthermore, fractions seem isolated from mathematics lessons and are taught like a separate topic.

TEST

A test was designed to answer our research questions. Its construction has been guided by our theoretical framework as well as the primary school curriculum in the French Community of Belgium. The test was split into two parts. Part A was made of 19 questions, Part B of 20 questions. There were 1 to 8 items for each question. There were 46 items in Part A and 48 in Part B. Part B was administered one week after Part A. Pupils had 50 min to answer each part.

Conceptual knowledge assessment

Conceptual knowledge of fractions was assessed through different categories of questions: part of a whole/partition, proportion and number. Three types of representations have been used: symbolic (e.g., 1/4), verbal (e.g., one-quarter) and figural representations (e.g., a square where the colored part represented 1/4). Discrete and continuous quantities were used.

Multiple variables were taken into account regarding numerical and verbal representations, such as the degree of familiarity, or the parity of the denominator and the numerator. The following variables were controlled regarding figural representations: the equivalence of the parts; the shape of the figure (square, rectangle, triangle . . .); the size of the figure; and the contiguity of the colored parts of the figure.

Part-whole/partition. Part-whole assessment included items for which children had to link fractions to a figural representation. The first question consisted of 6 items for which children were asked to represent a given fraction with a figure (e.g., draw a figure representing 1/7). The items were familiar fractions (1/2 and 3/4), unfamiliar fractions (1/7 and 4/5) and improper fractions (i.e., fractions larger than 1; 3/2 and 7/5). In the second question, pupils were asked to choose a figure representing a given fraction (e.g., choose figures representing 1/4, see Appendix). In the third question, they were asked to shade a certain portion of a figure. There were four items for this question. In the first two items, children were asked to shade 3/4 of a square or a rectangle. In the next two items, they were asked to shade 4/5 of a pentagon or a square.

Proportion. For questions about proportion, children were asked to compare quantities based on the rule of three. Five quantities were given in a table and they had to give the sixth quantity. There were verbal representations, such as "3 cakes cost €6, 5 cakes cost €10, 7 cakes cost €?" There were also figural representations. An example of figural representation is given in **Figure 1**. The contextualization of the items was introduced to make sure that children based their answer on both columns of the tables.

FIGURE 1 | Example of a figural proportion item.

Numbers. For the number category, there were four types of questions. The first question was a comparison of fractions. Pupils had to decide which of two fractions represented the larger quantity. There were fractions with the same numerator (e.g., 2/3_2/7), fractions with the same denominator (e.g., 3/8_5/8) and fractions with no common components (e.g., 2/5_1/4). In the second question, pupils were asked put fractions in ascending order. This question also involved improper fractions and natural numbers. The given numbers were the following: 3/4, 1/2, 8/4, and 1. The third question involved finding a fraction between two given fractions (e.g., find a fraction between 2/7 and 5/7). Fractions with common denominators, common numerators, and no common components were included. For the fourth question, pupils were asked to place a fraction or the unit on a graduated number line (e.g., given 0 and 1/4, place 3/4 on the number line). The given references were always 0 and another fraction.

Procedural knowledge assessment

We assessed the following procedures: addition and subtraction with or without the same denominator; multiplication of fractions; multiplication of a fraction by an integer; and simplification of fractions. Those procedures were assessed with typical questions such as $1/2 + 1/4 = ?$. Division of fractions was not included as it is not part of the official curriculum.

RESULTS

GENERAL RESULTS

Descriptive statistics are reported for each category of fractions (part-whole, proportion, numbers, operations, and simplification). Mean scores and standard deviations are always expressed in percentage. As can be seen in **Table 1**, children performed better for questions about proportion and part-whole than for questions about the other categories. There were still major difficulties in Grade 6 for the part-whole category. Indeed, even in

Grade 6, the percentage of correct responses was still far from ceiling performance. Children were capable of resolving questions on proportional reasoning from Grade 4. The main observed errors were linked to additive reasoning. Children got the lower scores in Grade 4 for arithmetic operations. This was not surprising as learning about operations on fractions usually start in Grade 5.

A correlation analysis was run to assess the relations between conceptual (part of a whole, proportion and numbers) and procedural categories (operations and simplification). The correlation analysis revealed that conceptual categories correlated significantly with each other (see **Table 2**). They also correlated positively with procedural categories.

We ran an ANOVA for repeated measures with category as a within-subjects factor (part-whole; proportion; number; operations; simplification) and grade as a between-subjects factor. There was a significant grade effect, $F_{(2, 437)} = 71.53$, $p < 0.001$, $\eta_p^2 = 0.25$. There was also a main effect of category, $F_{(4, 1744)} = 242.64$, $p < 0.001$, $\eta_p^2 = 0.36$, and a significant grade x category interaction, $F_{(8, 1744)} = 19.85$, $p < 0.001$, $\eta_p^2 = 0.08$ (see **Figure 2A**). Tukey *post-hoc* tests showed that accuracy for operations and simplification was poorer in Grade 4 than in Grades 5 and 6 ($p < 0.001$).

We ran another ANOVA for repeated measures on the type of knowledge (conceptual and procedural) with grade as a between-subjects factor. There was a significant effect of grade, $F_{(2, 437)} = 75.23$, $p < 0.001$, $\eta_p^2 = 0.26$. There was also a significant effect of the type of knowledge, $F_{(1, 438)} = 459.5$, $p < 0.001$, $\eta_p^2 = 0.51$, and a significant grade x type of knowledge interaction, $F_{(2, 437)} = 242.64$, $p < 0.001$, $\eta_p^2 = 0.36$ (see **Figure 2B**). Tukey *post-hoc* test was used to determine significant differences between grade mean values for each type of knowledge, revealing that performance was poorer for procedural knowledge in Grade 4 than in Grades 5 and 6 ($p < 0.001$).

We also ran cluster analyses to ensure that our categories reflected conceptual and procedural knowledge. Since two patterns appeared in the results, we ran two separate cluster analyses: one analysis for Grade 4 and one analysis for Grades 5 and 6. We ran neighbor-joining analyses (single linkage method) to see if our categories formed natural clusters that could be labeled according to a type of knowledge. These analyses provide a tree-structured graph (i.e., dendrogram) that is used to visualize the results of hierarchical clustering calculations. The dendrogram indicates at what level of similarity any two clusters were joined. It was constructed using neighbor-joining algorithm based on

Table 1 | Mean percentage of correct responses and standard deviation for each category in Grade 4–6.

	Part-whole	Proportion	Number	Operations	Simplification
Grade 4	65 ± 16	69 ± 28	47 ± 19	22 ± 18	26 ± 6
Grade 5	72 ± 13	78 ± 26	52 ± 18	37 ± 28	61 ± 9
Grade 6	77 ± 15	85 ± 22	63 ± 20	53 ± 27	71 ± 10

Table 2 | Correlations between conceptual items and procedural items.

	Part-whole	Proportion	Number	Operations	Simplification
Part-whole	1				
Proportion	0.348**	1			
Numbers	0.382**	0.359**	1		
Operations	0.383**	0.307**	0.460**	1	
Simplification	0.305**	0.386**	0.281**	0.387**	1

**Significant at p < 0.01.
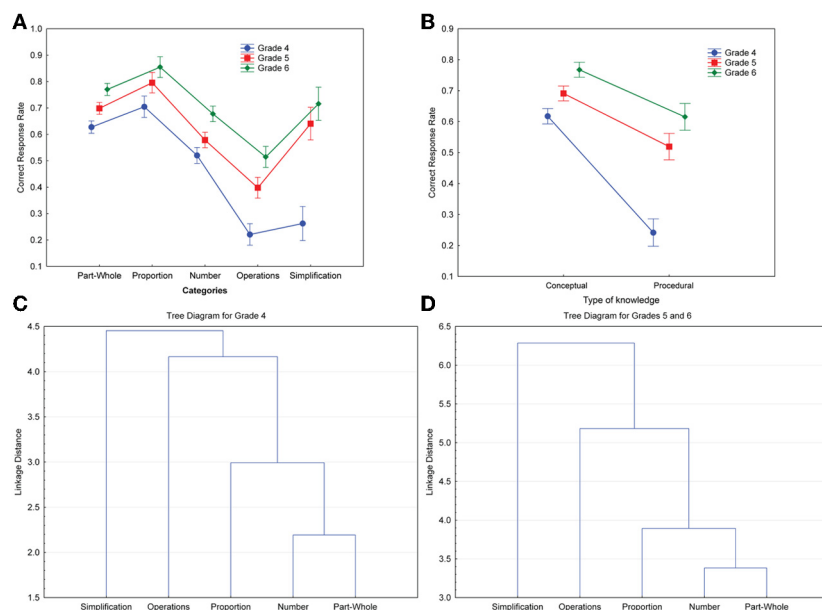
FIGURE 2 | The top two panels show the interaction between grade and correct response rates for each category (A), and between grade and each type of knowledge (B). Vertical bars denote 95% confidence intervals. The bottom two panels show dendrograms depicting the results of a single linkage hierarchical clustering of each category based on Euclidian distances for Grade 4 **(C)** and Grades 5 and 6 **(D)**.

Euclidian distances. Both for Grade 4 and for Grades 5 and 6, the dendrograms clustered the categories into two distinct groups that correspond to our two types of knowledge, i.e., conceptual and procedural (see **Figures 2C,D**). Part-whole, number and proportion were the most similar and correspond to our conceptual categories, whereas operations and simplification can be combined in a different cluster, that is our procedural categories.

PART-WHOLE/PARTITION

Draw a representation for each given fraction

Table 3 shows mean scores and standard deviation for the first question related to the part-whole/partition meaning of fractions. Different variables were involved in this question. Firstly, an ANOVA with the type of fraction as within-subject factor (2 levels: proper fraction vs. improper fraction) was run. Performance was worse for improper fractions than for proper fractions, $F_{(1, 438)} = 2039.2$, $p < 0.001$, $\eta_p^2 = 0.90$. Secondly, familiar (1/2, 3/4) and unfamiliar fractions (1/7, 4/5) were compared in another ANOVA. Performance for familiar fractions was significantly better than for unfamiliar fractions, $F_{(1, 438)} = 2406.9$, $p < 0.001$, $\eta_p^2 = 0.92$.

Despite potential graphic difficulties, pupils mostly divided a common continuous shape (circle or square, see **Figure 3**). 90% of pupils represented continuous quantities.

Select the figures representing 1/4

In this task, pupils had to choose figures representing the quantity 1/4 (see Appendix). Mean percentage of correct responses were high in every grade (Mean $= 92\% \pm 6\%$). But when figures were representing 2/8, we observed a dramatic drop of performance: $24 \pm 6\%$ in Grade 4, $29 \pm 8\%$ in Grade 5 and $59 \pm 9\%$ in

Table 3 | Mean percentage and standard deviation for the question: Draw a representation of the given fraction.

Items	Grade 4	Grade 5	Grade 6
1/2	84 ± 4	95 ± 14	98 ± 10
1/7	67 ± 7	83 ± 5	89 ±2
3/4	75 ± 8	87 ± 4	89 ± 3
4/5	67 ± 5	77 ± 3	90 ± 3
7/5	14 ± 7	20 ± 9	35 ± 9
3/2	23 ± 7	23 ± 11	41 ± 10

Grade 6. There was a significant difference between continuous and discrete quantities, $F_{(1, 438)} = 2308.1$, $p < 0.001$, $\eta_p^2 = 0.91$. Performance was better for continuous quantities.

Shade a certain fraction of a figure

In this task, pupils had to shade 3/4 or 4/5 of a given figure. Mean scores per grade are given in **Table 4**. Mean scores for 3/4 (Mean $= 83 \pm 2\%$) were higher than for 4/5 (Mean $= 65 \pm 4\%$). An ANOVA with familiarity as a within-subject factor showed a significant difference between 3/4 and 4/5, $F_{(1, 438)} = 3156.6$, $p < 0.001$, $\eta_p^2 = 0.93$.

PROPORTION

As seen in **Table 1**, performance for proportion items was better than in other categories. However, 10% of the answers given by 4th-graders were based on additive reasoning. This percentage dropped to 5% in Grade 5 and 2.6% in Grade 6. This type of error was more present for numerical items (Grade 4 = 9%; Grade 5 = 7%; Grade 6 = 3%) than for figural items (Grade 4 = 2%; Grade

5 = 2%; Grade 6 = 1%). A single-factor ANOVA was run and showed no significant difference between numerical and figural items, $F_{(1, 438)} = 0.6$, $p = 0.8$.

NUMBER

Place a given fraction on a number line
Percentage of correct responses showed a clear difference between three groups of items. In the first group of items, there were 3 number lines for which pupils only had to count the number of graduations corresponding to numerators to succeed (e.g., knowing 0 and 5/9 on the fifth graduation, place 2/9). For these items,

FIGURE 3 | Illustration of the most common answer when pupils were asked to draw a representation of a given fraction. 90% of them drew continuous quantities such as a circle or a rectangle. In this particular example, only 1/2 was represented correctly **(A)**. Parts of the drawings were unequal for 1/7 and 2/6 (**B** and **C**). Different shapes were used for 3/2 **(D)**.

Table 4 | Mean scores and standard deviation for each item in which pupils had to shade 3/4 or 4/5 of a given figure.

Figure	Fraction	Grade 4	Grade 5	Grade 6
▪	3/4	89 ± 2	88 ± 2	92 ± 1
▬	3/4	58 ± 5	82 ± 2	86 ± 1
▪	4/5	66 ± 4	62 ± 5	86 ± 2
⬠	4/5	55 ± 5	53 ± 5	69 ± 2

they could only process the numerator and ignore the denominator. Mean percentage of correct responses for these items was 89 ± 6%. In the second group of items, there were two number lines on which pupils had to place 1 (e.g., knowing 0 and 1/5 on the first graduation, place 1). The mean score for this group of items was the following: Mean = 40 ± 22%. The third group of items involved equivalent fractions (e.g., knowing 0 and 1/6 on the second graduation, place 2/3). The mean score for these items was the following: Mean = 31 ± 24%. An ANOVA with the group of items as a within-subject factor showed a significant difference between the first group of items compared to unit items and items involving equivalent fractions, $F_{(2, 437)} = 2942.6$, $p < 0.001$, $\eta_p^2 = 0.95$. Tukey *post-hoc* tests showed that the first group of items was higher than unit items ($p < 0.001$) and equivalent fractions items ($p < 0.001$).

Error analysis showed that when asked to place 1 on a number line, pupils had a tendency to place it at the beginning (12% of given responses) or at the end of the line (43% of given responses).

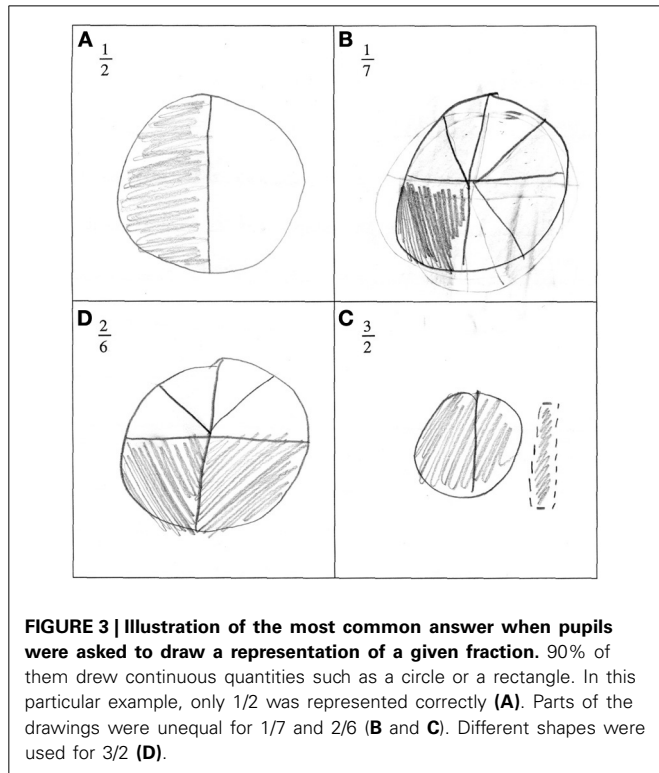### Put these fractions in ascending order
Children were asked to sort the following numbers in ascending order: 3/4, 1/2, 8/4, and 1. 55% of 4th-graders placed 1 at the end of the sequence, after 8/4. Furthermore, 22% of 4-graders placed 1 at the beginning of the sequence, before 1/2 and 3/4. This error rate decreased in grades 5 and 6, but 30% of 6th-graders still put 1 at the end of the sequence. These errors are consistent with the errors observed in the number line task. Children struggled with the relation between fractions and the unit.

Comparison of fractions
Pupils had to choose which of two fractions was larger. There were three types of items: same denominators (Mean = 83 ± 2%); same numerators (Mean = 56 ± 2%); and no common components (Mean = 65 ± 2%). An ANOVA on the type of fraction (3 levels: same denominators; same numerators; and no common components) revealed significant differences between types, $F_{(2, 437)} = 1346.4$, $p < 0.001$, $\eta_p^2 = 0.90$. Tukey *post-hoc* tests showed that scores for fractions with common denominators were higher than for fractions with common numerators ($p < 0.001$) and fractions with no common components ($p < 0.001$).

OPERATIONS
Performance for addition and subtraction with same denominators was better than for addition and subtraction with different denominators (see **Table 5**). This is not surprising as addition and subtraction with different denominators are not yet part of the program in Grade 4. But the procedure to find the lowest

Table 5 | Mean percentage of correct responses and standard deviation for each type of operations in Grade 4–6.

	Addition and subtraction/ same denominators	Addition and subtraction/ different denominators	Multiplication: fraction × integer	Multiplication: fraction × fraction
Grade 4	37 ± 9	1 ± 1	18 ± 7	39 ± 6
Grade 5	51 ± 7	25 ± 8	28 ± 5	36 ± 5
Grade 6	72 ± 7	33 ± 9	43 ± 5	54 ± 4

common denominator seems to pose problems in Grade 5 and 6. The most common error was based on the natural number bias, that is, adding or subtracting numerators and denominators as if there were natural numbers (e.g., = 1/3 + 1/4 = 2/7). 62% of 4th-graders made this mistake for addition and subtraction with different denominators, and this percentage still reached 22% in Grade 6. Surprisingly, performance for multiplication of fractions was better in Grade 4 than in Grade 5. An ANOVA showed significant differences on the types of operations, $F_{(2, 437)} = 135.5$, $p < 0.001$, $\eta_p^2 = 0.45$. Tukey *post-hoc* tests showed that performance was better for addition and subtraction with common denominators than for addition and subtraction with different denominators and multiplication ($p < 0.001$).

SIMPLIFICATION

As can be seen in **Table 6**, performance in the simplification task was better for fractions that could be divided by 2 (e.g., 4/8) than for fractions that could be divided by 3 (e.g., 15/9), $F_{(1, 438)} = 384.4$, $p < 0.001$, $\eta_p^2 = 0.64$. There was no significant difference between simplification of proper and improper fractions, fractions, $F_{(1, 438)} = 1.76$, $p = 0.19$.

DISCUSSION

In this study, we investigated the difficulties encountered by primary school children when learning fractions. One of the main goals of this study was to clarify the relationships between conceptual and procedural understanding of fractions. In order to do so, a test was administered in Grade 4–6 in classes of the French Community of Belgium. The test was based on the different conceptual meanings of fractions, namely part-whole/partition, number, proportion, as well as on procedural questions involving arithmetical operations and simplification of fractions.

Globally, the results showed large differences between categories. Pupils seemed to master the part-whole concept, whereas numbers and operations posed tremendous problems. Some conceptual meanings, such as numbers, were less used in primary school classes. Part-whole seems to be a concept that is widely used in the classrooms. Indeed, children performed well in the part-whole/partition category. However, they seem to have a stereotypic representation of fractions. Indeed, when they were asked to represent a given fraction, they mostly used a circle or a square, even when drawing collections could have been easier (e.g., 1/7). Moreover, when asked to select a figure representing a certain fraction, they performed better for continuous than discrete quantities. Pupils performed well with proportion items. These results contrast with textbooks and lessons given by teachers. In fact, the connection between proportions and fractions is

rarely made in textbooks and formal lessons, even if some aspects of fractions are based upon proportional reasoning (e.g., the rule of three).

In the proportion category, most errors were linked to additive reasoning. For example, when pupils are asked questions such as "3 cakes cost €12, 6 cakes cost €24, 8 cakes cost €?" the most common error would be the answer €36. In this case, children built their answer on only a subset of the given information and they applied additive strategies where multiplicative strategies should be used. Mistakes linked to additive reasoning are commonly reported during early stages of children's understanding of proportional reasoning (Lesh et al., 1988). This kind of mistakes was common in Grade 4, but could still be observed in Grade 6.

Pupils performed poorly in the numerical category. Even if children are trained to deal with number lines from grade 4, results showed major difficulties when they were asked to place a fraction on a graduated number line. They do not seem to have an appropriate representation of the quantities of fractions. Other studies have reported that many pupils experience difficulties when asked to locate a fraction on a number line. Pupils often view the whole number line, irrespective of its magnitude as a single unit instead of a scale (Ni, 2001). When they are asked to place a fraction between 0 and 1, pupils often place fractions disregarding any other reference point or known fractions. Pearn and Stephens (2004) pointed out that the incorrect location of fractions could also be the consequence of a lack of accuracy when dividing segments.

The lack of accuracy in children's mental representations of the magnitude of fractions seems to be confirmed by the weak percentage of correct response for questions involving sorting out a range of fractions in ascending order. Furthermore, mean percentage of correct responses for comparison of fractions were very low for fractions with common numerators and fractions no common components. When fractions share the same denominator (e.g., 2/5_4/5), the global magnitude of fractions is congruent with the magnitude of the numerators (e.g., 4 is larger than 2). In this case, pupils could only compare the numerators in order to choose the larger fraction. When fractions share the same numerator, the global magnitude of fractions is incongruent with the magnitude of denominators. Thus, pupils might not take the incongruity into account and their judgment might have been influenced by the whole number bias (Ni and Zhou, 2005). For fractions with no common components, pupils probably only compared numerators and denominators separately. This strategy led to larger error rates.

Focusing now on operations, children performed well in addition and subtraction of fractions with the same denominator, while performance dropped dramatically in addition and subtraction of fractions with different denominators. The most common errors were dictated by the whole number bias (Ni and Zhou, 2005). For example, when asked 3/4 + 2/5 = ?, the majority of pupils answers 5/9. Surprisingly, results were poorer for items involving the multiplication of an integer by a fraction, than for multiplication of two fractions. In the last case, pupils could successfully apply procedures based on natural numbers knowledge, which would explain higher percentage of correct response. Another surprising result was the better performance in Grade 4

Table 6 | Mean percentage of correct responses and standard deviation for the simplification task in each grade.

	Grade 4	Grade 5	Grade 6
4/10	36 ± 8	72 ± 11	78 ± 10
9/12	20 ± 6	54 ± 9	62 ± 11
15/9	19 ± 6	56 ± 9	71 ± 11
16/4	30 ± 6	63 ± 10	74 ± 9

than Grade 5 when children were asked to multiply an integer by a fraction. There might be a contamination of procedures applied to addition and subtraction with different denominators learnt in Grade 5.

Results showed massive familiarity effects in every category. Children performed significantly better on questions including familiar fractions, such as 1/2, 1/4, or 3/4 than on items with less familiar fractions. This could be due to the fact that the magnitude of 1/2 is known better than other fractional magnitudes. We do not know precisely when children start to quantify continuous quantities in informal contexts. Bryant (1974) suggests that children are able to understand part/part relations before part/whole relations. Relations such as "larger than/smaller than" and "equals to" could be the first logical relationships used at the beginning of fraction learning. Spinillo and Bryant (1991) designed experiments to analyse how 4- to 7-year-olds use the concept of "half" in equivalence judgment tasks. Their results suggest that using the concept of half would be the first step in relationships used by children to quantify fractions.

Desli (1999) also investigated the role of half by examining part/whole relationships. 6- to 8-year-olds were told that two parties had been organized and that chocolate bars would be equally distributed among children. They had to judge if they would receive the same amount of chocolate bars in both parties, and if not, in which party they would get more chocolate bars. Children had ceiling performance when they could use half as a reference. In the condition where they could not use half as a reference, only 8-year-olds had performance above chance. Desli (1999) also showed the importance of the concept of half in the construction of fractions quantifications. In a recent study using a fraction-based judgment task, Mazzocco et al. (2013) showed that fractions equivalent to 1/2 were easier to conceptualize. Moreover, children as young as 3 and 4 years old already have a good representation of the half boundary (Singer-Freeman and Goswami, 2001). As children are frequently exposed to 1/2 quite early in life, the familiarity of that quantity might induce a different type of mental representations compared to other less familiar fractions. Pupils might benefit from lessons including a larger pool of fractions. Teaching programs mostly insist on quantities that can be divided by 2. This limited vision of fractions seems to generate difficulties when it comes to generalization. Teachers could diversify the number of fractions used during lessons.

Improper fractions represented another major difficulty for primary school children (Bright et al., 1988; Tzur, 1999). The main difficulty appeared in the test when pupils were asked to graphically represent an improper fraction or when an improper fraction was presented in an ordering task. When pupils were asked to order 1 in a sequence involving fractions, the most common error was to put it at the end of the sequence, even if there was an improper fraction. This could mean that some children cannot imagine fractions can be larger than 1. This is consistent with the results found by Kallai and Tzelgov (2009) who showed that adults have a mental representation of what they called a "generalized fraction." A "generalized fraction corresponds to an "entity smaller than one" emerging from the common notation of fraction (Kallai and Tzelgov, 2009).

Furthermore, children seem to have a limited conception of the relation between 1 and fractions. Looking at questions on number lines and the ordering task, we observed two different conceptions regarding the number 1. In the first case, 1 was put at the beginning of the sequence. This can be interpreted as 1 being at the beginning of counting sequence. This error is again linked to the whole number bias (Ni and Zhou, 2005). Indeed, pupils based their answer on prior knowledge and the expectation that fractions follow the same rule of counting as whole numbers. In the second case, 1 was placed at the end of the sequence. Children who made this mistake considered fractions as being entities smaller than one.

Equivalent fractions were not understood by the majority of children (Kamii and Clark, 1995; Arnon et al., 2001). For example, performance was poor when they were asked to place 2/3 on a number line when the references were 0 and 1/6. Yet, their score was high for questions involving simplification of fraction. There was a clear dissociation between conceptual and procedural understanding. Children mastered the procedure applied to simplify fractions, but did not seem to understand the underlying concept of equivalent fractions.

To sum up, the test that we designed revealed many weaknesses in understanding fractions in primary school. Teaching practice seems to focus more on procedures than on conceptual understanding of fractions. But our results showed that procedures are not sufficient to carry out operations with fractions for instance. Even if pupils are intensively trained with finding the least common denominators procedure, the percentage of correct responses for addition and subtraction with different denominators remained low. Conceptual understanding is essential to ensure a deep understanding of fractions. In the U.S., it is already been recommend for the teaching of fractions (NCTM, 2000; Fazio and Siegler, 2012), and based on our results, we would suggest this recommendation should also apply for the French Community of Belgium.

We argue that children might benefit from a training based on concrete objects manipulation and explicit learning of rational numbers characteristics. Teaching children concrete activities could help them develop the corresponding abstract concepts (Arnon et al., 2001; Gabriel et al., 2012). For example, most primary school children consider fractions as being entities smaller than one (Behr et al., 1992; Stafylidou and Vosniadou, 2004). Moreover, most of them do not seem to understand equivalent fractions. These particular characteristics constitute the main differences between fractions and natural numbers. Pupils might benefit from more training with concrete objects to realize the necessary conceptual reorganisation and understand the properties of fractions. Another interesting finding of this study is that children performed better with familiar fractions. It could be interesting to introduce a larger variety as well as diversified representations of fractions in lessons. By integrating a larger range of fractions, children might get a more flexible representation of the magnitude of fractions.

Unfortunately, our experiment did not allow us to draw conclusions on how conceptual and procedural knowledge influence each other. Correlation analysis revealed that every conceptual and procedural items were positively correlated with each other.

Therefore, links between conceptual and procedural understanding are hard to interpret. This might mean that both types of knowledge are not independent and could be equally important when learning fractions. Both types of knowledge might evolve in an iterative way. Besides, individual differences have been reported in the development of conceptual and procedural knowledge (Hallett et al., 2010; Hecht and Vagi, 2012). Children differ in the use of conceptual and procedural knowledge to solve fraction problems (Hallett et al., 2010). Another reason can account for the difficulties to interpret findings obtained with a hypothetical measure of conceptual and procedural knowledge. The assessment of conceptual knowledge might reflect, to some extent, procedural knowledge and vice versa (Rittle-Johnson and Alibali, 1999). Future investigations are required to shed light on the links between conceptual and procedural knowledge in fraction learning and examine the possible reasons for individual differences.

In conclusion, our results showed that primary school children master the part-whole and proportion categories, but they struggle to understand fractions as numbers. Equivalent and improper fractions are very difficult to grasp, and pupils seem to apply procedures that they do not really understand. This might be linked to teaching practice that allocates more time and exercises only based on procedures.

ACKNOWLEDGMENTS

REFERENCES

Arnon, I., Nesher, P., and Nirenburg, R. (2001). Where do fractions encounter their equivalents? Can this encounter take place in elementary-school? *Int. J. Comput. Math. Learn.* 6, 167–214. doi: 10.1023/A:1017998922475

Baroody, A. J. (2003). "The development of adaptive expertise and flexibility: the integration of conceptual and procedural knowledge," in *The Development of Arithmetic Conceptsand Skills: Constructing Adaptive ExpertiseI*, eds A. J. Baroody and A. Dowker (Mahwah, NJ: Erlbaum), 1–33.

Behr, M. J., Harel, G., Post, T., and Lesh, R. (1992). "Rational number, ratio, and proportion," in *Handbook of Research on Mathematics Teaching and Learning*, ed D. A. Grouws (New York, NY: Macmillan), 296–333.

Behr, M. J., Lesh, R., Post, T. R., and Silver, E. A. (1983). "Rational numbers concepts," in *Acquisition of Mathematics Concepts and Processes*, eds R. Lesh and M. Landau (New York, NY: Academic Press), 91–125.

Bonato, M., Fabbri, S., Umiltà, C., and Zorzi, M. (2007). The mental representation of numerical fractions: real or integer? *J. Exp. Psychol. Hum. Percept. Perform.* 33, 1410–1419. doi: 10.1037/0096-1523.33.6.1410

Bright, G., Behr, M., Post, T., and Wachsmuth, I. (1988). Identifying fractions on number lines, *J. Res. Math. Educ.* 19, 215–232. doi: 10.2307/749066

Brissiaud, R. (1998). "Les fractions et les décimaux au CM1. Une nouvelle approche," in *Actes du XXVème Colloque des Formateurs et Professeurs de Mathématiques chargés de la Formation des Maîtres*, (IREM de Brest), 147–171.

Brousseau, G., Brousseau, N., and Warfield, V. (2004). Rationals and decimals as required in the school curriculum. Part 1: rationals as measurements. *J. Math. Behav.* 23, 1–20. doi: 10.1016/j.jmathb.2003. 12.001

Bryant, P. (1974). *Perception and Understanding in Young Children: An Experimental Approach*, Vol. 588. London: Methuen.

Byrnes, J. P., and Wasik, B. A. (1991). Role of conceptual knowledge in mathematical procedural learning. *Dev. Psychol.* 27, 777–786. doi: 10.1037/0012-1649.27.5.777

Charalambous, C., and Pitta-Pantazi, D. (2007). Drawing on a theoritical model to study students' understandings of fractions. *Educ. Stud. Math.* 64, 293–316. doi: 10.1007/s10649-006-9036-2

Coquin-Viennot, D., and Camos, V. (2006). "Décimaux et fractions," in *La cognition mathématique chez l'enfant*, eds P. Barrouillet and V. Camos (Marseille: Solal), 145–154.

Desli, D. (1999). *Children's Understanding of Intensive Quantities.* Unpublished PhD thesis. Institute of Education, University of London.

Fazio, L., and Siegler, R. (2012). Teaching fractions. *Educ. Pract. Ser.* 22, 1–28.

Gabriel, F., Coché, F., Szucs, D., Carette, V., Rey, B., and Content, A. (2012). Developing children's understanding of fractions: an intervention study. *Mind Brain Educ.* 6, 137–146. doi: 10.1111/j.1751-228X.2012. 01149.x

Gelman, R., and Williams, E. M. (1997). "Enabling constraints for cognitive development and learning: domain specificity and epigenesist," in *Cognitive Development, Handbook of Child Psychology, 5th Edn*, eds D. Kuhn and R. Siegler (New York, NY: Wiley), 575–630.

Grégoire, J. (2008). *Aux Sources des Difficultés de L'apprentissage des Fractions.* Brussels: Seminar given at, Université Libre de Bruxelles.

Grégoire, J., and Meert, G. (2005) "L'apprentissage des nombres rationnels et ses obstacles," in *Les Troubles Du Calcul*, ed M.-P. Noël (Marseille: Solal), 223–251.

Halford, G. S. (1993). *Children's Understanding: The Development of Mental Models*. Hillsdale, NJ: Erlbaum.

Hallett, D., Nunes, T., and Bryant, P. (2010). Individual differences in conceptual and procedural knowledge when learning fractions. *J. Educ. Psychol.* 102, 395–406. doi: 10.1037/a0017486

Hecht, S. A., and Vagi, K. J. (2012). Patterns of strengths and weaknesses in children's knowledge about fractions. *J. Exp. Child Psychol.* 111, 212–229.

Hecht, S. A., Close, L., and Santisi, M. (2003). Sources of individual differences in fraction skills. *J. Exp. Child Psychol.* 86, 277–302. doi: 10.1016/j.jecp.2003.08.003

Hiebert, J. (ed.). (1986). *Conceptual and Procedural Knowledge: the Case of Mathematics*. Hillsdale, NJ: Erlbaum.

Hiebert, J., and Lefevre, P. (1986). "Conceptual and procedural knowledge in mathematics: an introductory analysis," in *Conceptualand Procedural Knowledge: the Case of Mathematics*, ed J. Hiebert (Hillsdale, NJ: Lawrence Erlbaum Associates), 1–27.

Kallai, A., and Tzelgov, J. (2009). A generalized fraction: an entity smaller than one on the mental number line. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1845–1864. doi: 10.1037/a0016892

Kamii, C., and Clark, F. B. (1995). Equivalent fractions: their difficulty and educational implications. *J. Math. Behav.* 14, 365–378. doi: 10.1016/0732-3123(95)90035-7

Kerslake, D. (1986). *Fractions: Children's Strategies and Errors. A Report of the Strategies and Errors in Secondary Mathematics Project*. England: NFER-NELSON Publishing Company, Ltd.

Kieren, T. E. (1976). "On the mathematical, cognitive, and instructional foundations of rational numbers," in *Number and Measurement: Papers from a Research Workshop*, ed R. Lesh (Columbus, OH: ERIC/SMEAC), 101–144.

Kieren, T. E. (1988) "Personal knowledge of rational numbers: its intuitive and formal development," in *Research Agenda for Mathematics Education: Number Concepts and Operations in the Middle Grades*, Vol. 2, eds J. Hiebert and M. Behr (Virginia: Lawrence Erlbaum), 162–181.

Kieren, T. E. (1993). "Rational and Fractional Numbers: From Quotient fields to Recursive understanding," in *Rational Numbers: An Integration of Research*, eds T. P. Carpenter, E. Fennema, and T. A. Romberg (Hillsdale, NJ: Erlbaum), 49–84.

Lesh, R., Post, T., and Behr, M. (1988). "Proportional reasoning," in

Number Concepts and Operations in the Middle Grades, eds J. Hiebert and M. Behr (Reston, VA: Lawrence Erlbaum and National Council of Teachers of Mathematics), 93–118.

Mamede, E., Nunes, T., and Bryant, P. (2005). "The equivalence and ordering of fractions in part-whole and quotient situations," in *PME CONFERENCE*, Vol. 29, 3.

Mazzocco, M. M. M., Myers, G. F., Lewis, K. E., Hanich, L. B., and Murphy, M. M. (2013). Limited knowledge of fraction representations differentiates middle school students with mathematics learning disability (dyscalculia) vs. low mathematics achievement. *J. Exp. Child Psychol.* 115, 371–387. doi: 10.1016/j.jecp.2013.01.005

Ministère de la Communauté française, (1999). *Socles de Compétences Enseignement Fondamental et Premier degré de l'enseignement Secondaire. Test de Mathématique, Résultats et Commentaires.* Bruxelles: Administration générale de l'Enseignement et de la Recherche scientifique.

Moss, J., and Case, R. (1999). Developing children's understanding of the rational numbers: a new model and an experimental curriculum. *J. Res. Math. Educ.* 30, 122–147. doi: 10.2307/749607

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics.* Reston, VA: author.

Ni, Y. (2000). How valid is it to use number lines to measure children's conceptual knowledge about

rational number. *Educ. Psychol.* 20, 139–152. doi: 10.1080/713663716

Ni, Y. (2001). Semantic domains of rational numbers and the acquisition of fraction equivalence. *Contemp. Educ. Psycol.* 26, 400–417. doi: 10.1006/ceps.2000.1072

Ni, Y., and Zhou, Y. D. (2005). Teaching and learning fraction and rational numbers: the origins and implications of whole number bias. *Educ. Psychologist* 40, 27–52. doi: 10.1207/s15326985ep4001_3

Nunes, T., and Bryant, P. (1996). *Children Doing Mathematics.* Oxford: Blackwell.

Pearn, C., and Stephens, M. (2004). "Why you have to probe to discover what year 8 students really think about fractions," in *Paper Presented at the Mathematics Education for the Third Millennium: Towards 2010*, (Townsville).

Pitkethly, A., and Hunting, R. (1996). A review of recent research in the area of initial fraction concepts. *Educ. Stud. Math.* 30, 5–38. doi: 10.1007/BF00163751

Rittle-Johnson, B., Siegler, R. S., and Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: an iterative process. *J. Educ. Psychol.* 93, 346–362. doi: 10.1037/0022-0663.93.2.346

Rittle-Johnson, B., and Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: does one lead to the other? *J. Educ. Psychol.* 91, 175–189. doi: 10.1037/0022-0663.91.1.175

Rouche, N. (1998). *L'esprit des Sciences. Pourquoi ont-ils Inventé les Fractions?* Paris: Ellipses.

Schneider, M., and Stern, E. (2005). "Conceptual and procedural knowledge of a mathematics problem: their measurement and their causal interrelations," in *Paper Presented at the 27th Annual Meeting of the Cognitive Science Society (CSS)*, (Stresa).

Schneider, M., and Stern, E. (2010). The developmental relations between conceptual and procedural knowledge: a multimethod approach. *Dev. Psychol.* 46, 178–192. doi: 10.1037/a0016701

Siegler, R. S., Thompson, C. A., and Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cogn. Psychol.* 62, 273–296. doi: 10.1016/j.cogpsych.2011.03.001

Singer-Freeman, K. E., and Goswami, U. (2001). Does half a pizza equal half a box of chocolates?: proportional matching in an analogy task. *Cogn. Dev.* 16, 811–829. doi: 10.1016/S0885-2014(01)00066-1

Spinillo, A. G., and Bryant, P. (1991). Children's proportional judgments: The importance of "half". *Child Dev.* 62, 427–440. doi: 10.1111/j.1467-8624.1991.tb01542.x

Stafylidou, S., and Vosniadou, S. (2004). The development of students' understanding of the numerical value of fractions. *Learn. Instr.* 14, 503–518. doi: 10.1016/j.learninstruc.2004.06.015

Tzur, R. (1999). An integrated study of children's construction of improper

fractions and the teacher's role in promoting that learning. *J. Res. Math. Educ.* 30, 390–416. doi: 10.2307/749707

Vamvakoussi, X., and Vosniadou, S. (2004). Understanding the structure of the set of rational numbers: a conceptual change approach, *Learn. Instr.* 14, 453–467. doi: 10.1016/j.learninstruc.2004.06.013

APPENDIX

FIGURE A1 | Part-whole item: select fractions representing the same fraction as 1/4. To get full marks, children had to select panels **(A,B,D,E,G,H)** and **(J)**. In this particular example, the pupil failed to spot fractions equivalence.

The conceptual/procedural distinction belongs to strategies, not tasks: A comment on Gabriel et al. (2013)

Thomas J. Faulkenberry *

Department of Psychology and Counseling, Tarleton State University, Stephenville, TX, USA
Correspondence: faulkenberry@tarleton.edu

A commentary on

A componental view of children's difficulties in learning fractions

by Gabriel, F., Coché, F., Szucs, D., Carette, V., Rey, B., and Content, A. (2013). Front. Psychol. 4:715. doi: 10.3389/fpsyg.2013.00715

In their recent article, Gabriel et al. (2013) propose that the difficulty experienced by Belgian children in learning fractions stems from the fundamental dichotomy of procedural and conceptual knowledge of fractions. Indeed, the authors are not alone in this conclusion, as the procedural/conceptual knowledge divide has been a focus in several recent studies in numerical cognition (Hallett et al., 2010; Hecht and Vagi, 2012). Like their predecessors, Gabriel et al. adopt the definitions introduced by Rittle-Johnson and Alibali (1999), where conceptual knowledge refers to the understanding of the principles that govern a knowledge domain and procedural knowledge refers to knowledge of specific actions that are used to solve problems. This work stems from a rich foundation in mathematics education regarding the roles of procedural and conceptual knowledge in school-age children's mathematical development (Hiebert, 1986).

The issue I wish to raise in this commentary is not with the conclusions of Gabriel et al. (2013), but rather a general issue concerning the notion of conceptual and procedural *tasks* in mathematics. One of the difficulties is that a single task can reflect both types of knowledge. Indeed, consider the task of shading a geometric object (such as a square) to reflect a given fraction, say 3/4. This task appears from the outset to reflect conceptual knowledge, specifically since successful completion of the task gives some indication of the participant's knowledge of part-whole relationships in fractions. But, consider the alternative where a child is explicitly taught to perform this shading by first breaking the square into four equal sections, then shading three of the sections. Does this constitute a demonstration of conceptual knowledge? Or, does it reflect the use procedural knowledge? Hallett et al. (2010) would side with procedural knowledge since the procedure was explicitly taught to the child beforehand. Other authors (Hecht and Vagi, 2012; Gabriel et al., 2013) used this task and chose to call it a conceptual task. I would argue that the main question should not be whether the *task* is procedural or conceptual, but instead whether the employed *strategy* reflects the use of procedural or conceptual knowledge.

A related issue arises in Figure 3d of Gabriel et al. (2013). In this figure, a child has represented the fraction 2/6 by shading two parts of a circle that has been divided into six sections. The problem is that the two shaded sections are much larger than the four non-shaded sections, and as a result, the shaded fraction is actually 1/2. While this is a common error (and is apparently cross-cultural), it raises an important question. Does this child reflect conceptual understanding of fractions? I would argue no, since a conceptual understanding of the part-whole relationship would include the knowledge that the six parts should be *equal* in size. Of course, others may interpret this differently, and that is fine. My larger point is that it *can* be argued either way, and as such, the task does not define the knowledge that is used. Rather, it is the strategy used to complete the task that helps us make our conclusions.

Another common example of a task that can reflect both procedural and conceptual knowledge is the magnitude comparison task (e.g., which fraction is larger: 1/3 or 3/5?). Both Gabriel et al. (2013) and Hecht and Vagi (2012) termed this a conceptual task, presumably because it reflects the concept of fraction as a number. Indeed, if participants are forming mental representations of the fractions' magnitudes, then I feel that this is likely an accurate description. However, in a study with adults, Faulkenberry and Pierce (2011) found that on approximately 25% of trials, participants used a strategy known as cross-multiplication, where the size judgement is made by comparing cross-products (numerator of one fraction multiplied by denominator of another). This is a common procedural strategy that is employed in US schools for teaching students how to compare fractions (Boston et al., 2003). More importantly, this procedure allows the participant to arrive at an answer with no sense of the fraction as a number (i.e., no conceptual knowledge). Without explicitly asking participants to describe their solution strategy, it would not have been clear that they were using such a strategy, and by implication, it would have been impossible to know that they were using a *procedural* strategy on a task that looked conceptual.

From here, it is clear that there is a fundamental inconsistency in the literature. I think the proactive solution to this inconsistency lies in clearly delineating between the notions of conceptual/procedural

tasks and conceptual/procedural *strategies*. Simply coining tasks as procedural or conceptual does not appear to be sufficient. Rather, we need some knowledge of a participant's solution strategy to accurately determine which type of knowledge is responsible for the solution. Admittedly, this can be difficult to ascertain, as it requires interviewing participants about their methods of solving problems. While a trial-by-trial report of strategies may be the gold standard in this type of research, such data is time-consuming to gather. More optimistically, it may be possible to get a decent measure of strategy to simply asking participants *post-hoc* to simply describe how they solve problems of a given type. At the very least, tasks that are used in studies of conceptual/procedural knowledge should be subjected to at least two rounds of independent ratings of to how they reflect one type of knowledge or the other. This was the approach used in Hallett et al. (2010), and I feel that such data should be minimally required in future studies of this type.

In summary, I believe Gabriel et al. (2013) have conducted an important study in the field of numerical cognition of

fractions, particularly from the standpoint that it (1) identifies an important shortcoming of early fraction knowledge that appears to be cross-cultural, and it (2) begins an important dialogue about the methodological issues that we should consider when investigating the nature of conceptual and procedural knowledge in mathematics. We should continue to devote our time to serious investigations of the factors that influence conceptual and procedural knowledge in mathematics. At the same time, we should acknowledge that our current notion of labeling *tasks* as procedural or conceptual is limited, and that in the future we should investigate whether *strategies* employed on these tasks better reflect the use of procedural or conceptual knowledge.

REFERENCES

Boston, M. D., Smith, M. S., and Hillen, A. F. (2003). Building on students' intuitive strategies to make sense of cross multiplication. *Math. Teach. Mid. School* 9, 150–155.

Faulkenberry, T. J., and Pierce, B. H. (2011). Mental representations in fraction comparison: holistic versus component-based strategies. *Exp. Psychol.* 58, 480–489. doi: 10.1027/1618-3169/a000116

Gabriel, F., Coché, F., Szucs, D., Carette, V., Rey, B., and Content, A. (2013). A componential view of

children's difficulties in learning fractions. *Front. Psychol.* 4:715. doi: 10.3389/fpsyg.2013.00715

Hallett, D., Nunes, T., and Bryant, P. (2010). Individual differences in conceptual and procedural knowledge when learning fractions. *J. Educ. Psychol.* 102, 395–406. doi: 10.1037/a0017486

Hecht, S. A., and Vagi, K. J. (2012). Patterns of strengths and weaknesses in children's knowledge about fractions. *J. Exp. Child Psychol.* 111, 212–229. doi: 10.1016/j.jecp.2011.08.012

Hiebert, J. E. (1986). *Conceptual and Procedural knowledge: The Case of Mathematics.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Rittle-Johnson, B., and Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: does one lead to the other? *J. Educ. Psychol.* 91, 175–189. doi: 10.1037/0022-0663.91.1.175

Sixty-four or four-and-sixty? The influence of language and working memory on children's number transcoding

Ineke Imbo[1]*, Charlotte Vanden Bulcke[2], Jolien De Brauwer[3] and Wim Fias[1]

[1] Department of Experimental Psychology, Ghent University, Ghent, Belgium
[2] Department of Experimental-Clinical and Health Psychology, Ghent University, Ghent, Belgium
[3] Code, Expertise Centre for Development and Learning and Department of Applied Psychology, Thomas More University College, Antwerp, Belgium

Number transcoding (e.g., writing 64 when hearing "*sixty-four*") is a basic numerical skill; rather faultlessly performed in adults, but difficult for children. In the present study, children speaking Dutch (an inversed number language) and French (a non-inversed number language) wrote Arabic digits to dictation. We also tested their IQ and their phonological, visuospatial, and executive working memory. Although the number of transcoding errors (e.g., hearing 46 but writing 56) was equal in both groups, the number of inversion errors (e.g., hearing 46 but writing 64) was significantly higher in Dutch-speaking than in French-speaking children. Regression analyses confirmed that language was the only significant predictor of inversion errors. Working-memory components, in contrast, were the only significant predictors of transcoding errors. Executive resources were important in all children. Less-skilled transcoders also differed from more-skilled transcoders in that they used semantic rather than asemantic transcoding routes. Given the observed relation between number transcoding and mathematics grades, current findings may provide useful information for educational and clinical settings.

Keywords: number transcoding, number language, working memory, place-value understanding, transcoding errors, inversion errors

INTRODUCTION

Numeracy is extremely important in our everyday life (e.g., banking, cooking, shopping) and it gets even more important given the challenges of our modern society (e.g., population control, stock market crashes, climate change, and health risks, e.g., Reyna and Brainerd, 2007). Given the crucial role of numeracy, researchers started to search for its building blocks. The present study focuses on one of these building blocks, number transcoding or the ability to translate between number formats (verbal to Arabic or Arabic to verbal).

NUMBER TRANSCODING

Numbers come in various formats, such as Arabic digits (e.g., 46) and number words (e.g., "*forty-six*"). Number transcoding refers to the process in which a number is "translated" from one format into another one. Examples are writing down a dictated number or saying aloud an Arabic digit. Although these tasks are usually faultlessly performed by adults, they pose significant problems to young children. Throughout development, children learn to map the early-learned number words onto the respective Arabic symbols, increasing the overlap between the different formats (Kucian and Kaufmann, 2009).

The Arabic number system is rather simple, as it consists of only 10 elements (0, 1, 2, 3, 4, 5, 6, 7, 8, and 9) and one principle (i.e., the place value principle, according to which the value of a digit increases by a power of 10 with each step to the left). Verbal number systems, in contrast, are much more complicated. They rely on a limited lexicon, organized in different lexical classes such as units ("*one*" to "*nine*"), decades ("*ten*" to

"*ninety*"), hundreds, and thousands. Most lexicons also entail particulars such as "*eleven*" and "*twelve*." Because only a few quantities can be designated by a single word, a syntax provides the rules for making larger word sequences. Examples are additive rules (e.g., "*one hundred and sixty*" → 100 + 60 = 160) and multiplicative rules (e.g., "*four hundred*" → 4 × 100 = 400). Errors against these rules would result in 10060 and 4100, respectively.

Cognitive research into number transcoding was rather scarce, but gained renewed attention recently. The most widely used task to study number transcoding is writing Arabic numbers to dictation. Using this task, it has been shown that error rates are high, but decline across development: 49–54% in 6-year-olds, 22–54% in 7-year-olds, and 16–36% in 8-year-olds (Power and Dal Martello, 1990; Noël and Turconi, 1999; Camos, 2008; Zuber et al., 2009; Krinzinger et al., 2011; Pixner et al., 2011; Simmons et al., 2011). Syntactic errors, in which the number's elements are correct but its magnitude is not (e.g., "*one hundred twenty-three*" → 10023) are generally more frequent in children's performance than are lexical errors, in which a number's elements are incorrect (e.g., "*one hundred twenty-three*" → 124) (Seron et al., 1992; Sullivan et al., 1996).

Cognitive models of number transcoding offer a useful framework to understand and investigate number transcoding. A first category consists of *semantic* models, according to which the number word is first transformed into a semantic magnitude representation, and then into its constituent Arabic digits (e.g., McCloskey et al., 1985; Power and Dal Martello, 1990; McCloskey, 1992). Semantic models predict that number transcoding will be harder for larger numbers.

A second category consists of *asemantic* models, according to which number transcoding does not require any semantic magnitude representation (e.g., Deloche and Seron, 1987; Barrouillet et al., 2004). The ADAPT (A Developmental Asemantic Procedural Transcoding) model of Barrouillet et al. (2004), for example, consists of representational units in a mental lexicon, and production rules, according to which these units are combined. Since each part of the verbal number word is sent to working memory for storage and processing, the ADAPT model not only predicts that attentional resources are crucial in number transcoding, but also that number transcoding will be harder as the frequency of digits in the number increases.

A third category of models are connectionist models, which do not rely on rules or operators. The connectionist transcoding model of Verguts and Fias (2006) simulates transcoding on the basis of number frequency-based learning algorithms. Developed in the context of Arabic to verbal transcoding, the model generally demonstrates that a model without prior knowledge learns to read aloud numbers by developing two routes: a lexical route and a syntactic route. The lexical route is a direct route that maps Arabic input to phonological output representations without any intermediate steps. The syntactic route is an indirect route that applies principles of syntax to convert Arabic input to output phonology. These two routes do not involve any number magnitude representations, making the model essentially a non-semantic one. Yet, although not explicitly included in the model, Verguts and Fias (2006) do not exclude the possibility that an additional semantic route might exist for reading aloud a specific and restricted type of numbers, namely very small numbers and numbers with a very specific semantic meaning (like 1939). Their high frequency of occurrence (given their numerical magnitude, Dehaene and Mehler, 1992) and/or their special semantic status may make their semantic meaning more salient compared to other numbers with comparable numerical magnitude.

Whether children take a semantic or asemantic route depends on their numerical skill, as shown by Van Loosbroek et al. (2009). Nine-years old children with arithmetical disabilities needed more planning time writing large 1-digit numbers (e.g., 8) than when writing small 1-digit numbers (e.g., 3), suggesting the use of a semantic route. Control children, in contrast, did not show such a problem size effect for 1-digit numbers, suggesting an asemantic route. For 2- and 3-digit numbers, both groups of children showed a problem size effect, but the effect was smaller for control children.

LANGUAGE

Interestingly, recent research suggests that children's number transcoding is influenced by language. Writing numbers to dictation would be especially difficult for children speaking an inversed number language, such as Dutch or German, where the pronunciation of two-digit numbers is inversed (e.g., 64 is pronounced as "*four-and-sixty*"). In a cross-cultural transcoding study, Nuerk et al. (2005) observed that 7-year-old Japanese-speaking children made about six times fewer transcoding errors than did their German-speaking counterparts, and about eight times fewer inversion errors (e.g., hearing "*four-and-sixty*" but writing 46). Similarly, Krinzinger et al. (2011) showed that

Dutch- and German-speaking children made more transcoding errors than did French-speaking children. Pixner et al. (2011), finally, showed that 7-year-old children speaking Czech (which has an inverted and a non-inverted number language)[1] made 49% errors when numbers were dictated in the inverted number language, of which about half were inversion-related. When numbers were dictated in the non-inverted number language, errors dropped down to 37%.

One goal of the present study was to shed more light on these language effects in children's number transcoding. To that end, we tested children speaking an inversed number language (Dutch) and children speaking a non-inversed number language (French)[2]. Importantly, both groups were Belgian (Flemish and Walloons, respectively), so that cross-cultural differences in educational systems and math curricula were minimized. If the inversion property of the Dutch number language really affects children's number transcoding, more inversion errors should occur in the Dutch-speaking than in the French-speaking children. As neither semantic nor asemantic transcoding models do in their current form account for inversion errors, observing a reasonable number of inversion errors would urge for a revision of the transcoding models currently available.

However, it is interesting to examine not only the errors related to one specificity of a number language (such as inversion), but also the errors *not* related to this specificity. In this context, two accounts can be proposed (cf. Pixner et al., 2011). One possibility is that a complex number language puts a higher burden on children's available resources than a less complex number language. The complexity of an inverted number language would then also influence the execution of other (more general) transcoding rules, with more non-inversion errors in Dutch-speaking than in French-speaking children. In contrast, if the complexity of an inverted number language does not influence the execution of other transcoding rules, we would expect an similar number of non-inversion errors in Dutch-and French-speaking children.

WORKING MEMORY

Another goal of our study was to test the role of working memory in children's number transcoding. Camos (2008) showed that French 8-year-olds with lower working-memory spans made more transcoding errors than did 8-year-olds with higher working-memory spans. However, since Camos used a general working-memory span task (counting dots), it was impossible to pinpoint which working-memory component (phonological, visuospatial, or executive, following the three-component model of working memory; Baddeley and Hitch, 1974; Baddeley, 1992) was most important in children's number transcoding. Three recent studies (Zuber et al., 2009; Pixner et al., 2011;

---

[1] In Czech, the easier, non-inverted form of number words (e.g., 25 verbalized as "dvadsetpät" [twenty-five]) is taught at school, whereas the inverted form (e.g., 25 verbalized as "pätadvadset" [five-and-twenty]) is commonly used in daily life. There is thus still a confound between cultural and educational influences, on the one hand, and number language, on the other.

[2] In contrast to the French number language in France, where 70 is "soixante-dix" (sixty-ten) and 90 is "quatre-vingt-dix" (four-twenty-ten), the Walloons say "septante" and "nonante," respectively. All other numbers are pronounced similarly by Walloons and French children.

Simmons et al., 2011) extended Camos' (2008) research by including phonological, visuospatial, and executive tasks. As such, they could distinguish the impact of the different working-memory codes in number transcoding.

Zuber et al. (2009) showed that executive working memory was the strongest predictor of transcoding performance in German-speaking first graders: the higher a child scored on executive working-memory tasks, the fewer transcoding errors it committed. Further analyses showed that executive working memory was predictive of inversion-related errors whereas visuospatial working memory was predictive of non-inversion-related errors. The important role of the central executive was confirmed by Pixner et al. (2011). They showed that executive working memory predicted first graders' error rates in both inverted and non-inverted Czech number languages. Somewhat different results were obtained by Simmons et al. (2011). In their study on first and third graders speaking a non-inversed language (English), visuospatial working memory was the only significant predictor of transcoding performance. In contrast to Simmons et al.'s expectations, executive working memory was not predictive. Phonological working memory was predictive in none of the above-mentioned studies (Zuber et al., 2009; Pixner et al., 2011; Simmons et al., 2011), which is surprising because transcoding is commonly assumed to index verbal number processing.

It is, however, important to note that neither study tested the role of working memory across languages. Camos (2008) only tested French-speaking children, Zuber et al. (2009) only tested German-speaking children, Pixner et al. (2011) only tested Czech-speaking children, and Simmons et al. (2011) only tested English-speaking children. Although the results of these studies seem to suggest that executive working memory is more important in inverted number languages (Czech and German) than in non-inverted number languages (English), this conclusion might be premature. Indeed, the studies do not only differ in the languages they tested and in the tasks they used, but also in cultural and educational practices. In the present study, we tested both Dutch- and French-speaking children, which allowed us to test the role of working memory in inversed and non-inversed number languages, relatively independent of cultural and educational differences. If transcoding requires working-memory resources, working memory should predict both Dutch- and French-speaking children's transcoding performance. If transcoding is more resource-demanding in inversed than in non-inversed languages (Camos, 2008; Zuber et al., 2009; Pixner et al., 2011; Simmons et al., 2011), working memory should be more predictive in Dutch-speaking than in French-speaking children. However, if executing the inversion rule is what makes transcoding difficult, working memory should specifically predict Dutch children's number of inversion errors.

Another reason why Zuber et al. (2009) and Pixner et al. (2011) observed a role for executive working memory whereas Simmons et al. (2011) did not, may involve age. Zuber et al. (2009) and Pixner et al. (2011) tested first graders, who have no formal experience with numbers larger than 20, whereas Simmons and colleagues tested first and third graders, the latter having lots of formal experience with numbers larger than 20.

THE PRESENT STUDY

In sum, we wanted to test several hypotheses. First, which working-memory components are important in number transcoding? Second, what is the role of language in number transcoding? Do children speaking an inversed number language really make more transcoding errors than do children speaking a non-inversed number language? And if so, do children speaking an inversed number language rely more heavily on their working memory? Third, is the differentiation between semantic and asemantic transcoding routes (as observed by Van Loosbroek et al., 2009) also present in typically developing children?

Based on a pretest, we decided to focus on second graders, because first graders did not show enough knowledge of transcoding rules and because third graders did not make enough errors to allow a meaningful interpretation. Of the second graders we selected the 10 less- and more-skilled transcoders in each language group, which were further tested on IQ and working memory. Analyses on the percentages of transcoding and inversion errors are conducted, as well as analyses concerning the role of working memory. By dividing the children in less- and more-skilled transcoders, we were able to test whether the differentiation between semantic and asemantic transcoding routes (cf. Van Loosbroek et al., 2009) is also present in typically developing children. Since the asemantic route can be seen as developmentally more advanced than the semantic one (because there is no problem-size effect, the asemantic route can process more numbers in less time), we predicted that more-skilled children would use the asemantic route while less-skilled children would rather use the semantic route.

METHODS
PARTICIPANTS

A total of 87 children participated: 49 Dutch-speaking second graders (22 girls; mean age: 7 years 7 months) attending a school in the Flemish part of Belgium and 38 French-speaking second graders (20 girls; mean age: 7 years 7 months) attending a school in the Walloon part of Belgium. Mean age did not differ between both groups, $t_{(38)} = 0.00$; $p = 1.00$. Children only participated if they and their parents consented. None of the children presented sensory or motor deficiencies or any psychiatric diagnosis. The children received a small reward after participation. In each language group, children were ranked based on the total number of transcoding errors in number dictation. Ten children with the fewest and most errors were selected as the more-skilled and less-skilled transcoders group, respectively ($M_{transcoding\ errors} = 35.7$ for less-skilled Dutch-speaking children, $M_{transcoding\ errors} = 32.7$ for less-skilled French-speaking children, $M_{transcoding\ errors} = 0.30$ for more-skilled Dutch-speaking children, $M_{transcoding\ errors} = 0.7$ for more-skilled French-speaking children). None of these children were bilingual.

MATERIALS AND PROCEDURE

The number dictation task was presented to all children ($n = 87$), see below for a description. IQ and working memory were tested in the selected group only ($n = 40$), on two different days.

On the first day, working memory was tested by means of two phonological tasks (Digit and Letter span forward), two visuospatial tasks (Corsi blocks forward and Mazes memory), and four executive tasks (Digit and Letter span backward, Corsi blocks backward, and Sun moon Stroop). The Digit span, Corsi blocks, and Mazes memory tasks were taken from the Working Memory Test Battery for Children (WMTB-C, Pickering and Gathercole, 2001). On the second day, IQ was tested by means of two verbal subtests (Similarities and Vocabulary) and two performance subtests (Block design and Picture arrangement) of the WISC-III (Wechsler, 2002). These subtests provide a valid estimation of children's total IQ (Grégoire, 2001). Test-retest reliability is 0.92. The working-memory and IQ test series took about half an hour per child. If available, information about reliability of the measures is provided.

Number dictation task

The item set consisted of five 1-digit numbers, twenty 2-digit numbers, and forty 3-digit numbers. We made sure that each category of the ADAPT model was represented (see Supplementary material). The group-administered dictation took about 20 min and was conducted by the same, bilingual experimenter in both schools. The children received a booklet with 65 small pictures and were asked to write down the dictated number near the picture, mentioned as well during dictation (e.g., "write twenty-four next to the sun"). The pictures were used to motivate the children and to structure the responses of the dictation task. Each number was read aloud twice. When children did not know how to write the number, they were told to put an "X" instead.

Digit span forward and backward

The experimenter read a series of single-digit numbers at a rate of one digit per second, beginning with a string of 2 digits and proceeding to progressively larger strings, with a maximum of 9 digits. The child was required to repeat the exact sequence in the same (resp. reversed) order. There were six strings for each length, and testing was stopped when the child missed three sequences of the same length. Performance was scored as the number of correctly repeated digit strings. Test-retest reliability is 0.81 for digit span forward and 0.62 for digit span backward (WMTB-C, Pickering and Gathercole, 2001).

Letter span forward and backward

The method of this task is similar to the digit span, but the stimuli were letters instead of digits. Vowels and the letter w ("*double v*" in French) were not included, and all series consisted of phonologically different letters (cf. Butterworth et al., 1996). Performance was scored as the number of correctly repeated letter strings.

Corsi blocks forward and backward

The children were presented with nine identical wooden blocks in random positions on a wooden board. The children were told that these blocks were "stones in a pond." Using a plastic duck, the experimenter tapped on a sequence of blocks at the rate of one block per second, beginning with a 2-block sequence and proceeding to progressively larger sequences, with a maximum of 9 blocks. The child was asked to reproduce the exact sequence in the same (resp. reversed) order. There were six sequences

for each length, and testing was stopped when the child missed three sequences of the same length. Performance was scored as the number of correctly repeated sequences. Test-retest reliability for Corsi blocks forward is 0.53 (WMTB-C, Pickering and Gathercole, 2001).

Mazes memory

The child is presented with a picture of a maze, and a picture of an identical maze with the correct path drawn on it. The picture is removed, and the child's task was to duplicate the path in the response booklet. The difficulty level of the mazes started at span 2 (which corresponds to two walls in the maze), and proceeded to progressively larger spans, with a maximum of 8. At each level, the mazes get larger by one wall. There were six mazes for each level, and testing was stopped when the child missed three sequences of the same level. Performance was scored as the number of correctly drawn mazes.

Sun moon Stroop

This variant of the Stroop task is composed of two pages containing rows of pictures of suns and moons arranged pseudo-randomly (Archibald and Kerns, 1999). In the first condition, children are asked to say "sun" for a picture of a sun and "moon" for a picture of a moon. In the second condition, children were asked to say "sun" for a picture of a moon and "moon" for a picture of a sun. In both conditions, children were instructed to go as quickly and accurately as possible, within a time limit of 45 s. They had to stop and correct any errors that were made. If a child reached the end before the 45 s had elapsed, the time required to complete the page was recorded and the number that would have been correct within the time limit was estimated. A performance score was calculated by subtracting the number of correct responses in the first condition from the number of correct responses in the second condition and then dividing this difference by the number of correct responses in the first condition. Test-retest reliability is 0.86 (WMTB-C, Pickering and Gathercole, 2001).

RESULTS

ERROR ANALYSES

A categorization of all errors can be found in **Table 1**. Dutch- and French-speaking children made an equal number of transcoding errors [13% vs. 17%, $t_{(85)} = -1.05$, $p = 0.30$], see **Figure 1**. The percentage of transcoding errors was significantly higher on 3-digit than on 2-digit numbers, for both Dutch- and French-speaking children, $t_{(96)} = -4.37$ ($p < 0.001$) and $t_{(74)} = -4.53$ ($p < 0.001$), respectively. The percentage of inversions errors (among the total number of errors) was higher in Dutch-speaking than in French-speaking children (17 vs. 3%), $t_{(85)} = 3.56$ ($p < 0.001$), see **Figure 1**.

IQ AND WORKING MEMORY

Based on the number of transcoding errors, we selected the 10 more-skilled and 10 less-skilled transcoders in each language group. These children's IQ and working memory was further tested. Importantly, IQ scores differed neither between less- ($M = 106$) and more-skilled ($M = 112$) transcoders nor between Dutch- ($M = 106$) and French-speaking children ($M = 110$)
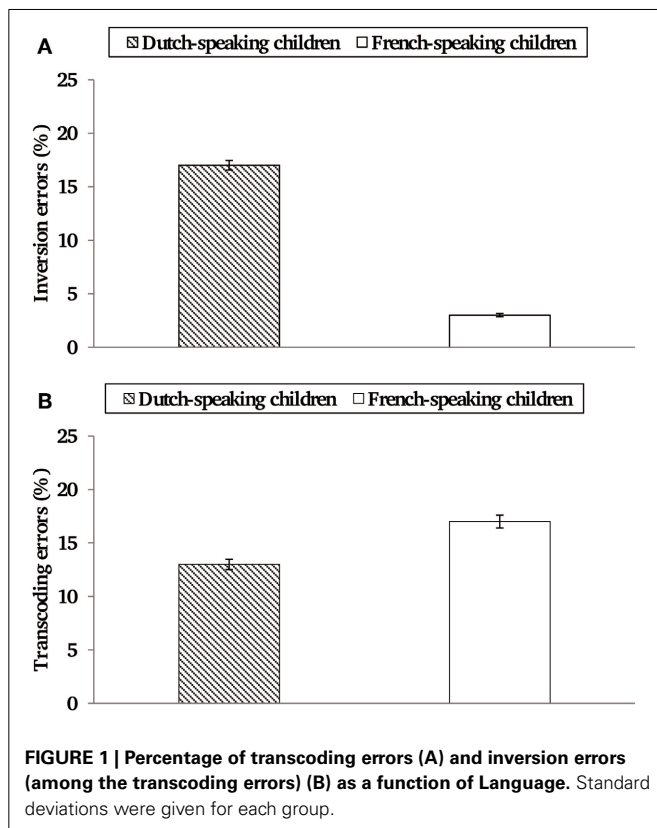
Table 1 | Items that were left open and erroneously transcoded as a function of Language (percentages between brackets).

	Dutch-speaking	French-speaking
Left open	$162 \pm 7 (5)$	$154 \pm 8 (6)$
Lexical error[a]	$36 \pm 1 (1)$	$52 \pm 2 (2)$
Syntactic error[b]	$293 \pm 10 (9)$	$192 \pm 10 (8)$
Combined error[c]	$90 \pm 5 (3)$	$189 \pm 9 (8)$

[a] Lexical error = when a lexical element is substituted by another one (e.g., $25 \rightarrow 24$).

[b] Syntactic error = when the elements of the number are correct but its magnitude is not (e.g., $123 \rightarrow 10023$).

[c] Combined error = when both lexical and syntactic rules are violated (e.g., $467 \rightarrow 40057$).

FIGURE 1 | Percentage of transcoding errors (A) and inversion errors (among the transcoding errors) (B) as a function of Language. Standard deviations were given for each group.

[respectively $t_{(38)} = -1.99$, $p = 0.06$; $t_{(38)} = -1.06$, $p = 0.29$]. The less- and more skilled transcoders differed in working-memory scores though. As can be seen in **Table 2**, the less-skilled transcoders scored lower on the Digit span forward, $t_{(38)} = 2.40$ ($p < 0.05$) and on the Letter span backward, $t_{(38)} = 3.39$ ($p < 0.01$). Correlations between IQ, age, and the working-memory tasks can be found in **Table 3**. To explore which working-memory components play a unique role in children's number transcoding, regression analyses were performed.

REGRESSION ANALYSES

The regression analyses incorporated compound scores for each working-memory component (phonological, visuospatial, and executive). These compound scores were calculated as the mean of the respective z-scores (for a similar procedure, see Barrouillet et al., 2008; Zuber et al., 2009; Pixner et al., 2011). Language (Dutch vs. French) was also included as a predictor, as were the interactions between language and the three working-memory components. Finally, IQ and age were included to ensure that potential working-memory influences were not due to IQ- or age-related differences.

In order to test the influence of language and working memory on children's transcoding performance, three regression analyses were performed. The predictors (phonological working memory, visuospatial working memory, executive working memory, language, IQ, age, and the interactions between language and the three working-memory components) were the same for the three regression analyses.

First, a binary logistic regression analysis was conducted with *type of transcoder* (less- vs. more-skilled) as dependent variable (A in **Table 4**). Executive working memory was the only significant predictor. Children with higher executive working-memory capacities had more chance to be more-skilled transcoders. A test of the final model (with executive working memory as a predictor) vs. the null model (with intercept only) was statistically significant, χ^2 (9, $n = 40$) = 14.58 ($p = 0.05$). The final model was able to correctly classify 68% of all children as being less- or more-skilled transcoders. Although digit span forward (a phonological working-memory task) differed between less- and more-skilled transcoders (**Table 2**), phonological working memory was not significantly predictive of type of transcoder (**Table 4**), probably because executive working memory accounted for most of the variance shared between these two working-memory components.

Second, a linear logistic regression analysis on the *number of transcoding errors* in less-skilled transcoders[3] (B in **Table 4**) shows that only phonological working memory tended to be a significant predictor ($p = 0.06$). Less-skilled transcoders with lower phonological working-memory scores made more transcoding errors ($R = 0.71$, adjusted $R^2 = 0.68$).

Finally, in a linear logistic regression analysis on the *number of inversion errors* (in all children), language was the only significant predictor (C in **Table 4**). Dutch-speaking children made more inversion errors than French-speaking children ($R = 0.76$, adjusted $R^2 = 0.55$).

ROLE OF SEMANTICS

In order to test whether children use a semantic or asemantic route when transcoding numbers, we explored the presence of problem size effects (cf. Van Loosbroek et al., 2009). Because the number of errors on 1-digit and 2-digit problems was very small, only 3-digit problems were included in this analysis. A median split was performed on 3-digit problems, dividing them in small problems ($M = 251$) and large problems ($M = 742$) with an equal number of transcoding rules [3.7 and 3.9, respectively, $t_{(38)} = -1.39$, $p = 0.17$]. Less-skilled transcoders made

---

[3]More-skilled transcoders were not included in this analysis because their number of transcoding errors was too small.

Table 2 | Scores on the working-memory tasks for more- and less-skilled transcoders.

	Phonological		Visuospatial		Executive			
	Digit forward	Letter forward	Corsi blocks	Mazes Memory	Digit backward	Letter backward	Corsi backward	Sun moon
More-skilled	27	23	19	13	12	11	14	14
Less-skilled	24	21	20	10	10	7	12	−11
t-test	−2.4*	1.78	0.46	1.29	−1.54	−3.39**	−1.23	1.24

*$p < 0.05$, **$p < 0.01$, (n = 20).

Table 3 | Correlations between IQ, age, and the working-memory tasks.

	1	2	3	4	5	6	7	8	9	10
1. IQ		−0.12	0.45**	0.22	0.24	0.11	−0.01	0.34*	0.15	0.10
2. Age			0.21	0.22	0.02	0.19	−0.16	0.05	0.16	0.04
3. Digit forward				0.38*	0.23	0.10	0.35*	0.55**	0.27	0.13
4. Letter forward					0.06	−0.10	0.07	0.03	0.01	−0.10
5. Corsi forward						0.02	0.18	0.14	0.29	0.22
6. Mazes memory							0.20	0.12	0.23	−0.06
7. Digit backward								0.64**	0.24	0.25
8. Letter backward									0.35*	0.21
9. Corsi backward										0.11
10. Sun moon										

*$p < 0.05$, **$p < 0.01$, (n = 40).

Table 4 | Standardized beta values of the three regression analyses.

	A (transcoder type)	B (transcoding errors)	C (inversion errors)
IQ	0.07	−0.04	0.00
Age	0.04	−0.16	0.03
Language	0.00	−0.55	0.51**
Visuospatial	−0.55	−2.73	0.09
Phonological	0.62	−3.95*	−0.38
Executive	1.67**	1.52	−0.67
Language × visuospatial	−0.07	−2.87	0.12
Language × phonological	−0.04	1.23	−0.54
Language × executive	0.35	−0.93	−0.19

* $p < .10$, ** $p < .05$, (n = 40).

significantly more transcoding errors on large 3-digit numbers than on small 3-digit numbers, $t_{(39)} = 26.19$ ($p < 0.001$), whereas there was no such difference in more-skilled transcoders, $t_{(39)} = 1.42$ ($p = 0.26$). Hence, less-skilled transcoders use a semantic route but more-skilled transcoders use an asemantic route.

MATHEMATICS ACHIEVEMENT

We compared the mathematics grades of the Dutch-speaking less- and more-skilled transcoders in our study. Mathematics grades (average grade in % for maths of the present school year) of the Dutch-speaking children were provided by the schools. Unfortunately, it was not possible to get mathematics

grades for the French-speaking children due to reasons of data protection. The more-skilled transcoders achieved significantly higher math scores (89%) than did the less-skilled transcoders (75%), $t_{(18)} = 3.15$ ($p < 0.01$).

DISCUSSION

We observed an equal number of transcoding errors in Dutch- and French-speaking children. Transcoding errors were more frequent on 3-digit than on 2-digit numbers, indicative of a role of working memory. Regression analyses confirmed that working memory played a significant role in number transcoding. The executive working memory component predicted whether children were less- or more-skilled transcoders. Interestingly, Dutch- and French-speaking children relied on executive resources to a similar degree (the executive working memory × language predictor was not significant). Regarding phonological working memory, we observed that less-skilled transcoders scored lower on the phonological working memory tasks compared to more-skilled transcoders. For the less-skilled transcoders phonological working memory turned out to be predictive of the number of transcoding errors (but the error rate of more-skilled transcoders was too low to test this in more-skilled transcoders). Visuospatial working memory was not predictive. The number of inversion errors in Dutch-speaking children was significantly higher than in French-speaking children. The regression analyses showed that language was the only significant predictor of inversion errors. Thus, although working memory plays an important role in transcoding in general it does not play a specific role in the application of the inversion principle.

LANGUAGE

Writing numbers to dictation is a task that adults perform rather faultlessly. Children, in contrast, experience many difficulties in this task (see **Table 1**). Second graders make fewer transcoding errors than do first graders, and third graders' transcoding performance is near to perfection. The number of transcoding errors did not differ between Dutch- and French-speaking children, indicating that the inversion property, specific for the Dutch number language, had no detrimental effect on children's general transcoding abilities. This is in contrast with the claim made by Pixner et al. (2011), who argued that the inversion property leads to a general increase in transcoding errors. However, in both our and Pixner et al.'s data, the number of non-inversion errors was actually *smaller* in the inversed number language than in the non-inversed number language, providing evidence against the claim that the inversion property would affect children's general transcoding abilities. It is clear that further research is needed into the occurrence of non-inversion errors. Which non-inversion errors are made, and are they more or less frequent in inverted number languages? In our data, for example, we noticed that French-speaking children made about 10% errors on numbers with 80, probably because of the complex French number word "*quatre-vingt*" [literally "*four-(times)-twenty*"].

Similar to transcoding errors, the number of inversion errors in the pretest decreased across age, with fewer inversion errors in second than in first graders, and no inversion errors in third graders. Dissimilar to transcoding errors, is that the number of inversion errors differed across Dutch- and French-speaking children. Dutch-speaking children made significantly more inversion errors than did French-speaking children. In fact, about 20% of the Dutch-speaking children's errors were inversion errors (see **Figure 1**). The inversion property of the Dutch number language thus results in committing specific errors (inversion errors) reflecting erroneous processing of the inversion rule. Regression analyses confirmed that the number of inversion errors was significantly predicted by a child's number language. These findings corroborate earlier findings (Nuerk et al., 2005; Krinzinger et al., 2011; Pixner et al., 2011; Simmons et al., 2011) by showing that a child's number language strongly influences its transcoding performance. Since in our study both Dutch- and French-speaking children attended Belgian schools, we can conclude that the language effects were truly linguistic, and could not be attributed to differential math curricula (see also Krinzinger et al., 2011).

Interestingly, inversion errors were not predicted by the interaction between language and any of the working-memory components. This indicates that children speaking an inversed number language rely as heavily on their working memory as do children speaking a non-inversed number language. Working memory did play a significant role in transcoding though (albeit the same role in inversed as in non-inversed languages), as discussed below.

WORKING MEMORY

Children made significantly more errors on 3-digit than on 2-digit numbers, as was predicted by the ADAPT model of Barrouillet et al. (2004). According to this model,

each part of the verbal number word is sent to working memory for storage and processing, which explains why number transcoding is harder as the number of digits increases.

Regression analyses were performed in order to test which working-memory components uniquely predicted children's transcoding performance. Executive working memory significantly predicted children's transcoding skill (i.e., less- vs. more-skilled). Children with more executive resources had more chance to be labeled as more-skilled transcoders. The significant role of the central executive is in agreement with Camos (2008), who observed that low-span children transcoded less efficiently than did high-span children, and with Zuber et al. (2009) and Pixner et al. (2011), who observed that executive working memory predicted children's number of transcoding errors. Transcoding verbal number words to Arabic symbols may rely on executive working memory (cf. attentional resources) for a high amount of processing steps, such as retrieving the respective Arabic number symbols from long-term memory, executing transcoding rules, resisting interference from incorrect number representations (e.g., hearing "*fifty*" but writing 15), and coordinating the retention of the partially completed digit chain while applying subsequent transcoding rules. The significant effect of executive working memory is in disagreement with Simmons et al. (2011), who—in contrast to their own expectation—observed no significant role for the central executive[4]. However, Simmons and colleagues note that the central executive not explaining unique variance does not exclude a role for this working-memory component. They contribute their null effect to the fact that the variance explained by the executive component covaried with the variance explained by the visuospatial component. In sum, our study confirms the conclusion, drawn by most recent studies as well, that executive resources play a significant role in children's number transcoding.

Our results may suggest that phonological working memory plays a role in transcoding as well, albeit only in less-skilled transcoders. Less-skilled transcoders with fewer phonological resources made more transcoding errors. Although more research is necessary, this is an important result, especially given the discrepancy between the predictions of the ADAPT model, according to which phonological resources are crucial in number transcoding (Barrouillet et al., 2004), and the absence of empirical evidence for this claim (Zuber et al., 2009; Pixner et al., 2011; Simmons et al., 2011). Phonological resources may be needed for the maintenance of intermediary information, such as the dictated verbal number word, the retrieved Arabic digits, and the chain under construction. Because less-skilled transcoders (who have fewer executive resources available) transcode less efficiently, they probably have more information to be stored simultaneously, increasing their reliance on phonological resources.

---

[4]Another peculiarity in Simmons et al. (2011) method is that the numbers were presented with increasing problem size. The children could thus reliably guess (a) the problem size of the next number, and (b) the number of digits in the next number, which might have decreased their need for executive control.

We did not observe a role of visuospatial working memory in number transcoding. This is in accordance with the results of Pixner et al. (2011), who neither observed a role for this working memory component in children's number transcoding. Our results are, however, in contrast with some other studies, where visuospatial working memory did predict children's transcoding errors (Simmons et al., 2011) or non-inversion errors (Zuber et al., 2009). Whether or not children retain a visuospatial representation of the digit chain when they are transcribing number words, is a question for further research. We propose that the answer to this question depends on several factors, such as the age of the children (with younger children relying exclusively on spatial coding and dual (visuospatial + phonological) coding arising around 8 years; Palmer, 2000; Pickering and Gathercole, 2001) and the type of errors (with visuospatial processes being more important in non-inversion errors, cf. Zuber et al., 2009).

Finally, it is important to note that none of the language × working memory variables did predict the number of transcoding errors. This indicates that the role of the different working-memory components is similar in children speaking inversed and non-inversed languages. This is somewhat surprising; given that the inversion property requires extra steps (such as memorizing and manipulating the sequence of number words), one might have expected that executive resources would be particularly related to inversion errors. As this study was the first one testing the role of working memory in children speaking inversed and non-inversed number languages, we hope that further studies will continue on this line of research. It might be interesting to contrast inversed number languages with completely transparent number languages (e.g., Chinese, where 264 is "*two-hundred-six-ten-four*").

IMPLICATIONS

Given the vast number of inversion errors, a first implication of our study is a theoretical one. As neither semantic nor asemantic transcoding models do in their current form account for inversion errors, this urges for a revision of these models. Note that the inversion principle does not only exist in Dutch, but also in other languages such as Arabic, Danish, Czech, German, Maltese, Malagasy, and Norwegian (Comrie, 2005). Hence, if transcoding models aim to be generally applicable, they have to add additional operators (such as inversion rules) in order to account for the errors observed in these languages. Connectionist modeling is informative here. Starting from a specific model like the one of Verguts and Fias (2006), it would of course be necessary to formally check if different network parameters, architectural constraints or training schemes would be necessary for a connectionist network to learn syntactic rules that include inversion. But, given that inversion by itself does not imply a serious rise of computational complexity or a change of underlying computational principle, there is no reason to expect that models of the type of Verguts and Fias (2006) are not capable of learning inversion. Hence, it can be reasonably expected that non-semantic models, equipped with a system for representing syntactic rules, can explain the specificities of inversion-related behavior.

A second theoretical implication concerns the role of working memory in number transcoding. The ADAPT model (Barrouillet et al., 2004) is the only transcoding model explicitly incorporating a role for working memory. As our and others' data show that working memory plays a significant role in children's number transcoding (Camos, 2008; Zuber et al., 2009; Pixner et al., 2011; Simmons et al., 2011), *all* transcoding models should actually pay attention to this influencing variable.

A last theoretical consideration concerns the dissociation between semantic and asemantic models: which type of models accounts best for the data? Based on recent evidence suggesting that route selection may depend on children's mathematical skill (with semantic routes being more frequently used in mathematically disabled children; Van Loosbroek et al., 2009), we tested if this was also true in typically developing children. According to our data, it looked as if less-skilled transcoders used a semantic route whereas more-skilled transcoders used an asemantic route. Hence, the selection between semantic and asemantic routes seems to depend on children's mathematical skill rather than being an all-or-none phenomenon. The observation of number size effects in the performance of mathematically disabled children might in principle reflect non-semantic effects of familiarity or of exposure rather than semantic effects emanating from the use of a semantic route. However, such an explanation is less likely to explain the number magnitude effect observed in the present study. Our participants were skilled readers and they had a highly similar educational curriculum. Therefore we can assume that they have had the same level of exposure to numbers. It would be interesting to further test this hypothesis not only as regards to error rates (as in the present study) but also as regards planning time (as in Van Loosbroek et al., 2009, who used graphic tablets that recorded children's pen trajectories). It would also be interesting to investigate the developmental trajectory to see whether the semantic transcoding established in less skilled readers presents as an intermediate stage in skilled transcoders at less skilled stages of development.

It should be noted that for cost-efficiency reasons, this study was conducted with only a small subset of less-and more skilled transcoders. Following Preacher et al. (2005), one should realize that this may come at a cost. First, it may lower power, especially if the two subsets are treated as a dichotomous variable. We do treat transcoding skill as a dichotomous variable in some analyses, like the t-tests and the binary regression. Yet, the fact that we also used regression analyses with number of errors as continuous dependent variable at least partially protects us against this problem. Second, using extreme groups leads to the inability to derive the exact nature of any non-linear relationship between the group variable and the variables under study. Because we consider this study to be exploratory and because existing theoretical models are not developed to such an extent that any claim is made about a precise relationship, we do not consider this to be a threat. Finally, one must be cautious in generalizing the results to all school children, given the small number of participants.

Besides the theoretical implications, there are also practical implications. Recent evidence suggests that number transcoding

may be a very important precursor of mathematical skill. In a longitudinal study, Moeller et al. (2011) showed that the number of inversion errors made in first grade reliably predicted children's addition performance in third grade. The number of inversion errors in first grade was also the only reliable predictor of mathematics grades in third grade. In the present study, we observed (for Dutch-speaking children only as mathematics grades were not available for French-speaking children) that the more-skilled transcoders achieved significantly higher math scores (89%) than did the less-skilled transcoders (75%), indicating that there is a meaningful relationship between the very basic skill of number transcoding and the more complex skill of mathematical problem solving. This implies that educators should give adequate attention to the mastery of the place-value system of the Arabic number system. Indeed, as argued by Moeller et al. (2011), transcoding errors indicate that children do not master the correspondence between verbal number words and the place-value structure for Arabic digits.

Given that transcoding errors can be an indication of arithmetical disabilities (e.g., Gross-Tsur et al., 1996; Hanich et al., 2001; Van Loosbroek et al., 2009), place-value understanding might also be a crucial factor in explaining mathematical difficulties. As such, we believe that further research should focus on number transcoding as an early precursor of later mathematical skill, and as a possible indicator of arithmetical disabilities.

AUTHOR NOTE

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fpsyg.2014.00313/abstract

REFERENCES

Archibald, S. J., and Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychol.* 5, 115–129. doi: 10.1076/chin.5.2.115.3167

Baddeley, A. (1992). Working memory. *Science*, 255, 556–559. doi: 10.1126/science.1736359

Baddeley, A. D., and Hitch, G. J. (1974). "Working memory," in *The Psychology of Learning and Motivation,* ed G. A. Bower (London: Academic Press), 47–89. doi: 10.1016/S0079-7421(08)60452-1

Barrouillet, P., Camos, V., Perruchet, P., and Seron, X. (2004). ADAPT: a developmental, asemantic, and procedural model for transcoding from verbal to Arabic numerals. *Psychol. Rev.* 111, 368–394. doi: 10.1037/0033-295X.111.2.368

Barrouillet, P., Mignon, M., and Thevenot, C. (2008). Strategies in subtraction problem solving in children. *J. Exp. Child Psychol.* 4, 233–251. doi: 10.1016/j.jecp.2007.12.001

Butterworth, B., Cipolotti, L., and Warrington, E. K. (1996). Short-term memory impairment and arithmetical ability. *Q. J. Exp. Psychol. A* 49, 251–262. doi: 10.1080/713755603

Camos, V. (2008). Low working memory capacity impedes both efficiency and learning of number transcoding in children. *J. Exp. Child Psychol.* 99, 37–57. doi: 10.1016/j.jecp.2007.06.006

Comrie, B. (2005). "Endangered numeral systems," in *Bedrohte Vielfalt: Aspekte des Sprach(en)tods [Endangered diversity: Aspects of Language Death],* eds J. Wohlgemuth and T. Dirksmeyer (Berlin: Weißensee Verlag), 203–230.

Dehaene, S., and Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition* 43, 1–29. doi: 10.1016/0010-0277(92)90030-L

Deloche, G., and Seron, X. (1987). "Numerical transcoding: a general production model," in *Mathematical Disabilities: A Cognitive Neuropsychological Perspective,* eds G. Deloche and X. Seron (Hillsdale, NJ: Erlbaum), 137–170.

Grégoire, J. (2001). Comparison of three short forms of the WISC-III. *Eur. Rev. Appl. Psychol.* 50, 437–441. doi: 10.1207/s15374424jccp2501_11

Gross-Tsur, V., Manor, O., and Shalev, R. S. (1996). Developmental dyscalculia: prevalence and demographic features. *Dev. Med. Child Neurol.* 38, 25–33. doi: 10.1111/j.1469-8749.1996.tb15029.x

Hanich, L. B., Jordan, N. C., Kaplan, D., and Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *J. Educ. Psychol.* 93, 615–626. doi: 10.1037/0022-0663.93.3.615

Krinzinger, H., Gregoire, J., Desoete, A., Kaufmann, L., Nuerk, H.-C., and Willmes, K. (2011). Differential language effects on numerical skills in second grade. *J. Cross Cult. Psychol.* 42, 614–629. doi: 10.1177/0022022111406252

Kucian, K., and Kaufmann, L. (2009). A developmental model of number representation. *Behav. Brain Sci.* 32, 340–331. doi: 10.1017/S0140525X09990069

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: evidence from acquired dyscalculia. *Cognition* 44, 107–157. doi: 10.1016/0010-0277(92)90052-J

McCloskey, M., Caramazza, A., and Basili, A. (1985). Cognitive mechanisms in number processing and calculation: evidence from dyscalculia. *Brain Cogn.* 4, 171–196 doi: 10.1016/0278-2626(85)90069-7

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., and Nuerk, H.-C. (2011). Early place-value understanding as a precursor for later arithmetic performance – a longitudinal study on numerical development. *Res. Dev. Disabil.* 32, 1837–1851. doi: 10.1016/j.ridd.2011.03.012

Noël, M.-P., and Turconi, E. (1999). Assessing number transcoding in children. *Eur. Rev. Appl. Psychol.* 49, 295–302.

Nuerk, H.-C., Olsen, N., and Willmes, K. (2005). "Better teach your children Japanese number words: how transparent number structure helps number acquisition," in *Poster Presented at the 23rd European Workshop of Cognitive Neuropsychology* (Bressanone).

Palmer, S. (2000). Working memory: a developmental study of phonological recoding. *Memory* 8, 179–193. doi: 10.1080/096582100387597

Pickering, S. J., and Gathercole, S. E. (2001). *Working Memory Test Battery for Children*. London: Psychological Corporation.

Pixner, S., Zuber, J., Hermanova, V., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). One language, two number-word systems and many problems: numerical cognition in the Czech language. *Res. Dev. Disabil.* 32, 2683–2689. doi: 10.1016/j.ridd.2011.06.004

Power, R. J. D., and Dal Martello, M. F. (1990). The dictation of Italian numerals. *Lang. Cogn. Process.* 5, 237–254. doi: 10.1080/01690969008402106

Preacher, K. J., Rucker, D. D., MacCallum, R. C., and Nicewander, W. A. (2005). Use of extreme groups approach: a critical reexamination and new recommendations. *Psychol. Methods* 10, 178–192. doi: 10.1037/1082-989X.10.2.178

Reyna, V. F., and Brainerd, C. J. (2007). The importance of mathematics in health and human judgment: numeracy, risk communication, and medical decision making. *Learn. Individ. Differ.* 17, 147–159. doi: 10.1016/j.lindif.2007.03.010

Seron, X., Deloche, G., and Noël, M. P. (1992). "Un transcodage des nombres chez l'enfant: la production des chiffres sous dictée," in *Les Chemins du Nombre,* eds J. Bideaud, C. Meljac, and J. P. Fisher (Lille: PUL), 303–327.

Simmons, F. R., Willis, C., and Adams, A.-M. (2011). Different components of working memory have different relationships with different mathematical skills. *J. Exp. Child Psychol.* 111, 139–155. doi: 10.1016/j.jecp.2011.08.011

Sullivan, K. S., Macaruso, P., and Sokol, S. M. (1996). Remediation of Arabic numeral processing in a case of developmental dyscalculia. *Neuropsychol. Rehabil.* 6, 27–53. doi: 10.1080/713755495

Van Loosbroek, E., Dirkx G. S. M. A., Hulstijn, W., and Janssen F. (2009). When the number line involves a delay: the writing of numbers by children of different arithmetical abilities. *J. Exp. Child Psychol.* 102, 26–39. doi: 10.1016/j.jecp.2008.07.003

Verguts, T., and Fias, W. (2006). Lexical and syntactic structures in a connectionist model of reading multi-digit numbers. *Conn. Sci.* 18, 265–283. doi: 10.1080/09540090600639396

Wechsler, D. (2002). *Wechsler Intelligence Scale for Children, 3rd Edn*. San Antonio, TX: Harcourt Assessment.

Zuber, J., Pixner, S., Moeller, K., and 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. *J. Exp. Child Psychol.* 102, 60–77. doi: 10.1016/j.jecp.2008.04.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Language influences on numerical development—Inversion effects on multi-digit number processing

E. Klein[1,2]*[†], J. Bahnmueller[3†], A. Mann[3], S. Pixner[4], L. Kaufmann[5], H.-C. Nuerk[2,3] and K. Moeller[2]

[1] Section Neuropsychology, Department of Neurology, University Hospital RWTH Aachen University, Tuebingen, Germany
[2] IWM-KMRC Knowledge Media Research Center, Tuebingen, Germany
[3] Department of Psychology, Eberhard Karls University, Tuebingen, Germany
[4] Department of Psychology, UMIT – The Health and Life Sciences University, Hall/Tyrol, Austria
[5] Department of Psychiatry and Psychotherapy A, General Hospital, Hall in Tyrol, Austria

In early numerical development, children have to become familiar with the Arabic number system and its place-value structure. The present review summarizes and discusses evidence for language influences on the acquisition of the highly transparent structuring principles of digital-Arabic digits by means of its moderation through the transparency of the respective language's number word system. In particular, the so-called inversion property (i.e., 24 named as "four and twenty" instead of "twenty four") was found to influence number processing in children not only in verbal but also in non-verbal numerical tasks. Additionally, there is first evidence suggesting that inversion-related difficulties may influence numerical processing longitudinally. Generally, language-specific influences in children's numerical development are most pronounced for multi-digit numbers. Yet, there is currently only one study on three-digit number processing for German-speaking children. A direct comparison of additional new data from Italian-speaking children further corroborates the assumption that language impacts on cognitive (number) processing as inversion-related interference was found most pronounced for German-speaking children. In sum, we conclude that numerical development may not be language-specific but seems to be moderated by language.

Keywords: number processing, numerical development, multi-digit number comparison, inversion effects, language-moderated effects

The Arabic number system is the world's most widely-used number system (see Zhang and Norman, 1995; Chrisomalis, 2004; Widom and Schlimm, 2012, for taxonomies of number systems). It relies on a simple formal structure: Based on a set of ten symbols (i.e., 0, 1, 2, 3, 4, 5, 6, 7, 8, 9) it is possible to assemble any possible number. All one has to consider is its *place-value structuring principle* which defines that the overall magnitude of a multi-digit number is coded by its constituting digits organized in descending powers of the basis 10 from left to right (i.e., $316 = \{3\} \times 10^2 + \{1\} \times 10^1 + \{6\} \times 10^0$). So, this principle defines the numerical value of each individual digit in a multi-digit number in digital-Arabic notation by its position within the respective digit string.

In early numerical development, children have to become familiar with Arabic numbers in general and, in particular, they have to understand the place-value principle with its underlying base-10-structure. However, this world-wide uniform and transparent combination principle only applies to numbers in digital-Arabic notation. In contrast, verbal number word systems differ between languages in the way they correspond to the systematic and language-independent place-value structuring of digital-Arabic numbers. This is often referred to as the transparency of a number word system, describing how closely a language's number word system conforms to the structure of a digital number system (Dowker et al., 2008). Importantly, there

is accumulating evidence that difficulties in understanding the Arabic number system and interferences on number processing are associated with the transparency of number word systems (e.g., Seron and Fayol, 1994; Nuerk et al., 2005a; Zuber et al., 2009; Pixner et al., 2011a). In particular, language (by means of its number word structure) moderates multi-digit number processing as indicated by a variety of language moderated effects in adults but also in children (see Nuerk et al., 2011 for a review). However, the question when and how these language modulations become important in numerical development is still not answered sufficiently.

Several studies showed that children speaking a language with transparent number words have fewer problems acquiring basic arithmetical competencies. In particular, most Asian languages (e.g., Japanese, Korean, Chinese) have a transparent number word system (Comrie, 2005). For instance, in Japanese 452 is literally coded as "four-hundred-five-ten-two" ("yon-hyaku-go-jū-ni"). In this vein, Miura et al. (1993) showed that Asian as compared to English-speaking first graders exhibited better understanding of the place-value structure of the Arabic number system when asked to, e.g., explicitly identify tens and units of a two-digit number. Based on such results, it has also been argued that the higher mathematical achievement observed repeatedly for Asian as compared to European and American children may benefit from their more transparent number word systems (Miura et al.,

1993, 1994; Towse and Saxton, 1998). This argument was corroborated by a recent systematic review by Ng and Rao (2010) indicating that these differences in mathematical achievement cannot be accounted for entirely by cultural influences (i.e., educational system, student motivation, etc.) but are—at least in early mathematical development—driven by the benefits of Asian number word systems.

This seems plausible when considering that the number word systems of most European languages are less transparent. For instance, there are specific words for multiples of ten (e.g., "forty" and not "four ten"), teen numbers (e.g., "sixteen" and not "ten six"), number words with a base other than ten (e.g., in French 82 is named as "quatre-vingt-deux" which translates to "four twenty two"). One of the most important inconsistencies in number words common to several languages (e.g., Arabic, Danish, German, Maltese, etc.) is the inversion property. In this context, inversion describes the fact that the order in which tens and units are named in number words is inverted compared to their order in digital-Arabic notation. For instance, in German 24 is named "vierundzwanzig" which literally translates to "fourandtwenty." Mastering this inconsistency poses one of the most common challenges in early numerical development for children speaking a language with inversion. Importantly, difficulties related to the inversion property are not restricted to transcoding and the use of number words but also generalize to the processing of number magnitude.

INVERSION EFFECTS ON TRANSCODING

Influences of the inversion property on verbal numerical tasks such as transcoding have been shown repeatedly. For instance, Krinzinger et al. (2011) evaluated transcoding performance of 2nd graders from France, Wallonia, Flanders, Germany, and Austria. Their results indicated inversion-based between-language effects when pupils had to write down Arabic numbers to dictation: those speaking a language with inverted number words (i.e., Flemish, Austrian, German) made generally more transcoding errors as those speaking a language without inversion. More particularly, Zuber et al. (2009) investigated inversion-related transcoding errors in German-speaking 1st graders and observed that almost half of their errors were related to the inversion property of German number words. In line with this, Nuerk et al. (2005a) observed that German-speaking 1st graders not only committed significantly more transcoding errors in general as compared to Japanese-speaking children but more inversion-related errors in particular.

Interestingly, there are two number-word systems in Czech—one with and one without inversion. Pixner et al. (2011b) observed that the same children committed more errors and, in particular, more inversion-related errors when asked to transcode number words dictated in their inverted compared to their non-inverted form. These results and the fact that inversion-related transcoding errors have not been reported for languages without inversion (French: Barrouillet et al., 2004; Camos, 2008; Italian: Power and Dal Martello, 1990, 1997) clearly suggest that intransparencies of a language's number word system such as the inversion property may impede the acquisition of basic numerical skills.

INVERSION EFFECTS ON THE PROCESSING OF NUMBER MAGNITUDE

Influences of inversion have also been observed for tasks with no explicit involvement of number words such as Arabic number magnitude comparison and number line estimation. As regards magnitude comparison, the so-called unit-decade compatibility effect (UDCE, Nuerk et al., 2001) was found to be moderated by inversion. The UDCE describes the finding that magnitude comparisons of unit-decade-compatible number pairs (e.g., 32–57, $3 < 5$ and $2 < 7$) are faster and less error-prone than comparisons of unit-decade-incompatible pairs (e.g., 37–62; $3 < 6$, but $7 > 2$). Thereby, the compatibility effect indicates influences of decision-irrelevant units on the overall comparison process. This suggests that the magnitude of units, tens, hundreds, etc. are also represented in a decomposed manner complying with the place-value structure of the Arabic number system (cf. Nuerk et al., 2011 for a review). Although the UDCE was observed for German-speakers first it is not specific for languages with inversion. It was observed in several other languages, both with inverted number words (Dutch: Ratinckx et al., 2006) and without inverted number words (English: Nuerk et al., 2005b; Moeller et al., 2009; Spanish: Macizo and Herrera, 2008; Italian: Macizo et al., 2010; Pixner et al., 2011a; Hebrew: Ganor-Stern et al., 2007, 2009). While the effect is not language-specific, it is, however, language-moderated. It was found to be more pronounced in languages with inversion both in children (Pixner et al., 2011a) and adults (Nuerk et al., 2005b). Interestingly, Pixner et al. (2011a) investigated the UDCE in German- (language with inverted number words), Italian- (without inversion) and Czech-speaking (both inverted and non-inverted number words) 1st graders. As indexed by the size of the UDCE, the interference due to decision irrelevant units was most pronounced for the language with inverted number words (German), followed by the language having both inverted and non-inverted number words (Czech) and the language without inversion (Italian).

Moreover, inversion-related language differences were also observed for the number line estimation task. Siegler and Mu (2008) showed that Chinese children's number line estimations were more accurate than those of North American children (see also Muldoon et al., 2011). Additionally, Helmreich et al. (2011) found that estimates were more accurate for Italian-speaking as compared to German-speaking children. While these language differences fit nicely with the fact that the Chinese number word system is more transparent than the English and the English more transparent than the German, one cannot exclude that the observed differences may also be driven by more general cultural differences (e.g., curricular differences). Therefore, it seems to be more promising to investigate possible influences of differences between number word systems more specifically with respect to the influence of the inversion property. In this context, Helmreich et al. (2011) identified two specific effects of inversion on children's number line estimations. First, the authors manipulated inter-digit distance of the to-be-estimated numbers [large, e.g., for 28 ($8 - 2 = 6$) vs. small, e.g., for 45 ($5 - 4 = 1$)]. Between-language differences should be more pronounced for large inter-digit distances, because marking 82 instead of 28 leads to a larger estimation error as compared to marking 45 instead

of 54. And indeed, the overall advantage in estimation accuracy of Italian-speaking children was driven by target numbers with a large inter-digit distance. Second, the resulting error bias should be systematic with respect to its direction. For numbers like 49, children should overestimate the position on the number systematically because 94 (when confusing tens and units) is larger than the correct target 49 and vice versa for numbers like 51 following the same rationale. Helmreich et al. (2011) observed that this directional bias was more pronounced for German-speaking than for Italian-speaking children. Thus, even though German-speaking children refer to the same underlying mental number line representation, they were hampered to integrate tens and units into a coherent representation of a two-digit number due to the inversion property of the German number word system.

Taken together, both the results for number magnitude comparison as well as number line estimation indicate that even in non-verbal numerical tasks the transparency of the respective language's number word system influences number processing skills in children. In particular, inversion-related intransparency caused systematic and significant performance shortcomings. This raises the question whether these influences are developmentally relevant.

LANGUAGE INFLUENCES ON NUMERICAL DEVELOPMENT?

Generally, associations between number word inversion and numerical performance are developmentally relevant when inversion-related shortcomings predict future numerical development. Importantly, Moeller et al. (2011) were able to show such a longitudinal influence for German-speaking children. Inversion errors in transcoding and the size of the compatibility effect in 1st grade, which are both more frequent/pronounced in languages with inverted number words and indicate early place-value understanding reliably predicted arithmetic performance in 3rd grade: the more inversion transcoding errors a child committed and the larger her/his compatibility effect in 1st grade the more errors a child made in an addition task two years later. Importantly, this association was reliable even after controlling for general cognitive ability and working memory. However, they also found more specific inversion-related effects: more inversion errors in 1st grade predicted a larger carry effect in addition as a criterion for place-value processing in 3rd grade. Finally, Moeller et al. (2011) also found that the percentage of inversion-related transcoding errors predicted the mathematics mark at the end of 3rd grade reliably: more inversion-related errors were associated with a worse mathematics mark. Importantly, this also indicates that deficiencies in early place-value understanding do not sort itself out over time.

Unfortunately, there is currently no study contrasting the influence of the inversion property on children's numerical development in a longitudinal and cross-cultural approach. However, above longitudinal influences of inversion-moderated effects for German-speaking children clearly suggest that number word structure moderates children's numerical development differentially. Fewer (or even no) inversion transcoding errors and a smaller UDCE—as also found for languages without inverted number words (see above)—were associated with better arithmetic performance.

Yet, to date the majority of research on inversion influences focuses on two-digit numbers. This seems obvious as it is the order of tens and units only that is inverted. However, one might expect inversion to also influence three-digit number processing. While in Italian or English, the neighboring number word constituents correspond to the neighboring Arabic digit, this is not the case in German: 329 is named "three-hundred-nine-and-twenty". Thus, the neighboring number words are "three" (for hundreds) and "nine" (for units), whereas the neighboring digits are "3" (for hundreds) and "2" (for tens).

LANGUAGE EFFECTS ON THREE-DIGIT NUMBER PROCESSING

Currently, only few studies extend the UDCE to three-digit numbers. For English-speaking adults, Korvorst and Damian (2008) suggested that place-value and single digit magnitude information is automatically taken into account when processing three-digit numbers. Hundred-decade and hundred-unit compatibility effects (HDCE and HUCE, respectively) indicated decomposed processing of units, tens, and hundreds. However, they also observed that the HUCE (see **Table 1**) was smaller than the HDCE and argued that units may cause less interference than tens because of a left-to–right processing gradient for multi-digit numbers (see also Poltrock and Schwartz, 1984).
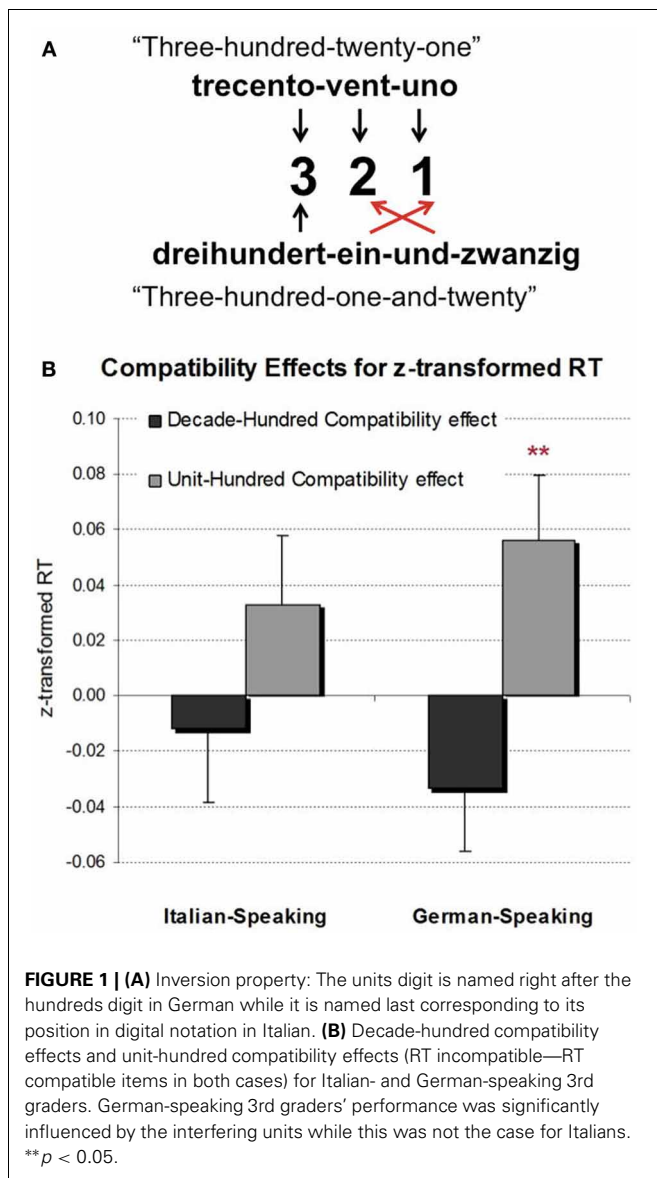
For children, Mann et al. (2012) investigated the HDCE and HUCE in German-speaking students longitudinally from grade two to four observing the HUCE to increase with age. However, the HDCE did not reach significance for any grade level. Importantly, the inversion property of the German number word system offers a plausible account for this pattern: because units are named directly after hundreds and thus before the tens (i.e., 239 → "zweihundertneununddreißig" meaning "two-hundred-nine-and-thirty," see **Figure 1A**) interference due to the unit-digit might be more pronounced than interference due to the tens-digit. Thus, interference by the neighboring number word constituents was more pronounced than by the neighboring Arabic digits.

However, because there are only these data from German-speaking children, conclusions about possible language differences have to be drawn cautiously as a direct contrast of compatibility effects for three-digit numbers between different language groups is still missing. To present a first perspective on

Table 1 | Example stimuli.

	Decade-hundred compatible		Decade-hundred incompatible	
Unit-hundred compatible	4 1 2 ∧ ∧ ∧ 9 5 7		8 3 6 ∨ ∧ ∨ 2 9 1	
Unit-hundred incompatible	8 7 1 ∨ ∨ ∧ 3 2 6		1 8 6 ∧ ∨ ∨ 7 3 1	

Place-value processing was indexed by compatibility effects (incompatible minus compatible), this means by the interference caused by the value of digits which were irrelevant for the overall magnitude decision because of their position in the digit string.

FIGURE 1 | (A) Inversion property: The units digit is named right after the hundreds digit in German while it is named last corresponding to its position in digital notation in Italian. **(B)** Decade-hundred compatibility effects and unit-hundred compatibility effects (RT incompatible—RT compatible items in both cases) for Italian- and German-speaking 3rd graders. German-speaking 3rd graders' performance was significantly influenced by the interfering units while this was not the case for Italians. **p < 0.05.

language differences for numbers beyond the two-digit number range, we briefly present additional new data on three-digit number comparison of 82 Italian-speaking 3rd graders (40 female; mean age 9;0 years; SD = 3.5 months; non-inverted number words) to contrast them with those of the German sample of Mann et al. (2012; 96 children, 47 female, mean age 9;4 years, SD = 4.4 months; inverted number words).

All participants completed a three-digit number magnitude comparison task. In the stimulus set of 80 between-hundred three-digit number pairs the factors decade-hundred compatibility and unit-hundred compatibility were manipulated orthogonally (see **Table 1**). Children had to indicate the larger of two simultaneously presented numbers by pressing a corresponding button. RT analyses were based exclusively on correct between-decade trials. Additionally, a trimming procedure first eliminated RT shorter than 200 ms and larger than 8000 ms and then all RT deviating from the individual's mean by more than 3

SD. As RT means and SD varied considerably between participants, RT were z-transformed (zRT) prior to the analyses. Please note, the pattern for error rates was similar ($r = 0.80$) but due to the generally low error rates ($M = 4.0\%$; $SD = 2.4\%$) less discriminating.

We observed a regular HUCE [$F_{(1, 176)} = 6.77$, $p < 0.05$] indicating that hundred-unit compatible items were responded to faster (1431 ms) than hundred-unit incompatible items (1450 ms). However, as the pattern of compatibility effects was similar in both language groups the interaction of language group and compatibility was not significant [$F_{(1, 176)} = 3.56$, $p = 0.56$]. Nevertheless, as we had a specific hypothesis regarding the presence of the compatibility effect, we inspected the simple effects. These indicated that the HUCE was only significant for German-speaking 3rd graders [$t_{(95)} = 2.50$, $p < 0.05$, see **Figure 1B**] with hundred-unit compatible items responded to faster (1258 ms) than hundred-unit incompatible items (1278 ms), but not for Italian-speaking 3rd graders [$t_{(81)} = 1.26$, $p = 0.21$].

These findings further corroborate the assumption that magnitude processing of multi-digit numbers and, in particular, the processing of place-value information is moderated by linguistic characteristics such as the inversion property of the respective number word structure. Unit interference was only significant for German-speaking 3rd graders. As previously found for two-digit numbers, these results suggest that proximity not only in digital but also in verbal number word notation is a relevant predictor for place-value compatibility effects. On a more general level, the data provide further support to the notion that language impacts on cognitive (number) processes, even those supposed to be non-verbal. Interestingly, the pattern of compatibility effects was similar for both languages. This fits nicely with the results of Helmreich et al. (2011, see above). In a number line estimation task, both estimation errors due to large inter-digit distances as well as the directional bias of estimation errors were more pronounced for German-speaking but nevertheless present for Italian-speaking children. However, while this influence of inversion on number line estimation was significant in direct comparisons (see also Nuerk et al., 2005a,b; Pixner et al., 2011a for evidence on magnitude comparison), it was not significant in the data presented here. Nevertheless, the simple effect analyses indicate that three-digit number processing also seems to be influenced by inversion—although to a weaker degree compared to two-digit numbers. Moreover, this suggests that other language invariant aspects moderate (multi-digit) number processing. As regards three-digit numbers, perceptual attributes such as lateral masking effects may be an influencing factor because hundred and unit digits flank the tens digit in the center from both sides, possibly overcoming (linguistic) effects for those digits.

CONCLUSION AND PERSPECTIVES

The current review indicates that numerical development may not be language-specific but is, however, moderated by language. Language influences in children's numerical development seem to be more pronounced for multi-digit numbers complying with the fact that differences between number word systems studied so far have been stronger for multi- than for single-digit numbers.

In particular, place-value integration is more difficult in languages with inverted number words in which units are named before tens. New data comparing German-speaking and Italian-speaking children in three-digit number comparison generalize this assumption beyond the two-digit number range. The findings discussed in this review are highly relevant for numerical development since inversion-related difficulties were shown to predict later arithmetic performance (Moeller et al., 2011).

However, apart from specificities of number word systems there are other more general language specificities which may influence number processing such as, for instance, reading direction (e.g., Shaki et al., 2009, see Göbel et al., 2011 for a review). These studies provide conclusive evidence that reading direction influences number processing in adults, in particular spatial numerical associations. Shaki et al. (2009) observed that English-speaking participants (who read both words and numbers from left to right) systematically associated small numbers with left and large numbers with right whereas this association was reversed for Palestinians (reading words and Arabic-Indic numbers from right to left). Against this background, it might be interesting to further investigate language influences by means of different number systems as well as reading direction to further evaluate the impact of language on numerical development.

Another interesting question regards the directionality of influences between number processing and language. To the best of our knowledge there is currently no study investigating possible influences of number processing on language. One possible reason for this uni-directional research bias might be that for an approach paralleling the one pursued in most of the studies described above one would need two cultures speaking the same language but having a different number system to investigate the influence of number processing on language. Unfortunately, we do not know any such case. Maybe it might be possible to address this issue in a less strict manner in languages such as Czech (with both inverted and non-inverted number words). So far, there is only evidence for specific difficulties in the number domain associated with the use of the inverted form (e.g., Pixner et al., 2011a,b). However, one might think of investigating how the use of either number word system influences language processing in these children in future studies.

REFERENCES

Barrouillet, P., Camos, V., Perruchet, P., and Seron, X. (2004). ADAPT: a developmental, asemantic, and procedural model for transcoding from verbal to Arabic numerals. *Psychol. Rev.* 111, 368–394. doi: 10.1037/0033-295X.111.2.368

Camos, V. (2008). Low working memory capacity impedes both efficiency and learning of number transcoding in children. *J. Exp. Child Psychol.* 99, 37–57. doi: 10.1016/j.jecp.2007.06.006

Chrisomalis, S. (2004). A cognitive typology for numerical notation. *Camb. Archaeol. J.* 14, 37–52. doi: 10.1017/S0959774304000034

Colomé, A., Laka, I., and Sebastiaen-Galles, N. (2010). Language effects in addition: how you say it counts. *Q. J. Exp. Psychol.* 63, 965–983. doi: 10.1080/17470210903134377

Comrie, B. (2005). "Endangered numeral systems," in *Endangered Diversity: Aspects of Language Death*, eds J. Wohlgemuth and T. Dirksmeyer (Berlin: Weißensee Verlag), 203–230.

Dowker, A., Bala, S., and Lloyd, D. (2008). Linguistic influences on mathematical development: how important is the transparency of the counting system? *Philos. Psychol.* 21, 523–538. doi: 10.1080/09515080802285511

Ganor-Stern, D., Pinhas, M., and Tzelgov, J. (2009). Comparing two-digit numbers: the importance of being presented together. *Q. J. Exp. Psychol.* 62, 444–452. doi: 10.1080/17470210802391631

Ganor-Stern, D., Tzelgov, J., and Ellenbogen, R. (2007). Automaticity of two-digit numbers. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 483–496. doi: 10.1037/0096-1523.33.2.483

Göbel, S. M., Shaki, S., and Fischer, M. H. (2011). The cultural number line: a review of cultural and linguistic influences on the development of number processing. *J. Cross Cult. Psychol.* 42, 543–565. doi: 10.1177/0022022111406251

Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). Language effects on children's mental number line: how cross-cultural differences in number word systems affect spatial mappings of numbers in a non-verbal task. *J. Cross Cult. Psychol.* 42, 598–613. doi: 10.1177/0022022111406026

Korvorst, M., and Damian, M. F. (2008). The differential influence of decades and units on multi-digit number comparison. *Q. J. Exp. Psychol.* 61, 1250–1264. doi: 10.1080/17470210701503286

Krinzinger, H., Gregoire, J., Desoete, A., Kaufmann, L. Nuerk, H.-C., and Willmes, K. (2011).Differential language effects on numerical skills in second grade. *J. Cross Cult. Psychol.* 42, 614–629. doi: 10.1177/0022022111406252

Macizo, P., and Herrera, A. (2008). El efecto del código numérico en la tarea de comparación de números de dos cifras. The effect of number codes in the comparison task of two-digit numbers. *Psicologica* 29, 1–34.

Macizo, P., Herrera, A., Paolieri, D., and Román, P. (2010). Is there cross-language modulation when bilinguals process number words? *Appl. Psycholinguist.* 31, 651–669. doi: 10.1080/S0142716410000184

Mann, A., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H.-C. (2012). On the development of Arabic three-digit number processing in primary school children. *J. Exp. Child Psychol.* 113, 594–601. doi: 10.1016/j.jecp.2012.08.002

Miura, I. T., Okamoto, Y., Kim, C. C., Chang, C.-M., Steere, M., and Fayol, M. (1994). Comparisons of children's cognitive representation of number: China, France, Japan, Korea, Sweden and the United States. *Int. J. Behav. Dev.* 17, 401–411. doi: 10.1177/016502549401700301

Miura, I. T., Okamoto, Y., Kim, C. C., Steere, M., and Fayol, M. (1993). First graders' cognitive representation of number and understanding of place value: cross-national comparisons – France, Japan, Korea, Sweden, and the United States. *J. Educ. Psychol.* 85, 24–30. doi: 10.1037/0022-0663.85.1.24

Moeller, K., Nuerk, H.-C., and Willmes, K. (2009). Internal magnitude representation is not holistic, either. *Eur. J. Cogn. Psychol.* 21, 672–685. doi: 10.1080/09541440802311899

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., and Nuerk, H.-C. (2011). Early place-value understanding as a precursor for later arithmetic performance – a longitudinal study on numerical development. *Res.*

Dev. Disabil. 32, 1837–1851. doi: 10.1016/j.ridd.2011.03.012

Muldoon, K., Simms, V., Towse, J., Menzies, V., and Yue, G. (2011). Cross- comparisons of 5-year-olds' estimating and mathematical ability. *J. Cross Cult. Psychol.* 42, 669–681. doi: 10.1177/0022022111406035

Ng, S. S. N., and Rao, N. (2010). Chinese number words, culture, and mathematics learning. *Rev. Educ. Res.* 80, 180–206. doi: 10.3102/0034654310364764

Nuerk, H.-C., Moeller, K., Klein, E., Willmes, K., and Fischer, M. H. (2011). Extending the mental number line - A review of multi-digit number processing. *J. Psychol.* 219, 3–22. doi: 10.1027/2151-2604/a000041

Nuerk, H.-C., Olsen, N., and Willmes, K. (2005a). "Better teach your children Japanese number words: how transparent number structure helps number acquisition," in *Poster presented at the 23rd Workshop of Cognitive Neuropsychology* (Bressanone).

Nuerk, H.-C., Weger, U., and Willmes, K. (2005b). Language effects in magnitude comparison: small, but not irrelevant. *Brain Lang.* 92, 262–277.

Nuerk, H.-C., Weger, U., and Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition* 82, B25–B33. doi: 10.1016/S0010-0277(01)00142-1

Pixner, S., Kaufmann, L., Moeller, K., Hermanova, V., and Nuerk, H.-C. (2011a). Whorf reloaded:

language effects on non-verbal number processing in 1st grade - a trilingual study. *J. Exp. Child Psychol.* 108, 371–382. doi: 10.1016/j.jecp.2010.09.002

Pixner, S., Zuber, J., Hermanova, V., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011b). One language, two number-word systems and many problems: Numerical cognition in the Czech language. *Res. Dev. Disabil.* 32, 2683–2689. doi: 10.1016/j.ridd.2011.06.004

Poltrock, S. E., and Schwartz, D. R. (1984). Comparative judgement of multi-digit numbers. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 32–45. doi: 10.1037/0278-7393.10.1.32

Power, R. J. D., and Dal Martello, M. F. (1990). The dictation of Italian numerals. *Lang. Cogn. Process.* 5, 237–254. doi: 10.1080/01690969008402106

Power, R. J. D., and Dal Martello, M. F. (1997). From 834 to eighty thirty four: the reading of Arabic numerals by seven year old children.

Math. Cogn. 3, 63–85. doi: 10.1080/135467997387489

Ratinckx, E., Nuerk, H.-C., van Dijk, J.-P., and Willmes, K. (2006). Effects of interhemispheric communication on two-digit number processing. *Cortex* 42, 1128–1137. doi: 10.1016/S0010-9452(08)70225-9

Seron, X., and Fayol, M. (1994). Number transcoding in children: a functional analysis. *Br. J. Dev. Psychol.* 12, 281–300. doi: 10.1111/j.2044-835X.1994.tb00635.x

Shaki, S., Fischer, M. H., and Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychon. Bull. Rev.* 16, 328–331. doi: 10.3758/PBR.16.2.328

Siegler, R. S., and Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychol. Sci.* 19, 759–763. doi: 10.1111/j.1467-9280.2008.02153.x

Towse, J. N., and Saxton, M. (1998). "Mathematics across national boundaries: cultural and linguistic

perspectives on numerical competence," in *The Development of Mathematics Skills,* ed C. Donlan (Hove: Psychology Press), 129–150.

Widom, T. R., and Schlimm, D. (2012). Methodological reflections on typologies for numerical notations. *Sci. Context* 25, 155–195. doi: 10.1017/S0269889712000038

Zhang, J., and Norman, D. A. (1995). A representational analysis of numeration systems. *Cognition* 57, 271–295. doi: 10.1016/0010-0277(95)00674-3

Zuber, J., Pixner, S., Moeller, K., and 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. *J. Exp. Child Psychol.* 102, 60–77. doi: 10.1016/j.jecp.2008.04.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships

that could be construed as a potential conflict of interest.

Phonemic awareness as a pathway to number transcoding

Júlia B. Lopes-Silva[1,2]*, Ricardo Moura[1,3], Annelise Júlio-Costa[1,3], Vitor G. Haase[1,2,3] and Guilherme Wood[4]

[1] Developmental Neuropsychology Laboratory, Department of Psychology, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[2] Programa de Pós-graduação em Saúde da Criança e do Adolescente, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[3] Programa de Pós-graduação em Neurociências, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[4] Department of Neuropsychology, Institute of Psychology, Karl-Franzens-University of Graz, Graz, Austria

Although verbal and numerical abilities have a well-established interaction, the impact of phonological processing on numeric abilities remains elusive. The aim of this study is to investigate the role of phonemic awareness in number processing and to explore its association with other functions such as working memory and magnitude processing. One hundred seventy-two children in 2nd grade to 4th grade were evaluated in terms of their intelligence, number transcoding, phonemic awareness, verbal and visuospatial working memory and number sense (non-symbolic magnitude comparison) performance. All of the children had normal intelligence. Among these measurements of magnitude processing, working memory and phonemic awareness, only the last was retained in regression and path models predicting transcoding ability. Phonemic awareness mediated the influence of verbal working memory on number transcoding. The evidence suggests that phonemic awareness significantly affects number transcoding. Such an association is robust and should be considered in cognitive models of both dyslexia and dyscalculia.

Keywords: phonemic awareness, verbal working memory, transcoding, ADAPT, asemantic transcoding models

INTRODUCTION

Mastering reading and writing numbers in their verbal and Arabic forms is an essential skill for daily life (Lochy and Censabella, 2005). Being able to manipulate numbers and convert them from one format into another is one of the first steps in children's mathematical learning and starts to be formally trained in kindergarten. The ability to establish a relationship between the verbal and Arabic representations of number, when a conversion of numerical symbols from one notation to the other is necessary, is called number transcoding (Deloche and Seron, 1987).

The verbal number system is linguistically structured and, although it may differ among languages, there are some common basic principles and regularities (Fayol and Seron, 2005). It is typically composed of a lexicon of single words that designate a few quantities (like *five*, *eleven*, *seventy* and *hundred*) and organized by a syntax that arranges these lexical units in order to represent any possible quantity. The two basic syntactic principles are the relations of addition and multiplication. In this sense, numbers are represented as sum relationships (e.g.: *eighty-one* means *eighty* plus *one*) and product relationships (e.g.: *three hundred* means *three times hundred*). The number words in Portuguese are similar to the English number words in the sense that they are also organized in lexical classes for units, decades and particulars (the *-teens* in English) (Wood et al., 2006).

The Arabic code is more complex and is acquired later in development (Geary, 2000). Its lexicon is composed of only a small set of different symbols (digits from 0 to 9), and the basic syntactic principle that combines them to form all numbers is the positional value (or place-value). According to this principle, the digit's value depends on its position in the numerical string and is given by a power of base ten. Therefore, in the case of three-digit numbers, the first digit (from right to left) is multiplied by 10^0, the second by 10^1, and so on. The number 124, for example, represents a quantity equal to $1 \times 10^2 + 2 \times 10^1 + 4 \times 10^0$ (or $100 + 20 + 4$). The digit 0 has a special syntactic role when it denotes the absence of a given power of ten, as occurs in numbers with internal zeros, for example the number 406 ($4 \times 10^2 + 0 \times 10^1 + 6 \times 10^0$).

One preeminent model of number transcoding is ADAPT (A Developmental, Asemantic, and Procedural model for Transcoding from verbal to Arabic numerals; Barrouillet et al., 2004). According to ADAPT, the inputs are coded into a phonological sequence and the parsing mechanisms then subdivide this sequence into smaller units to be processed by a production system. This production system is related to rules devoted to the retrieval of Arabic forms from long-term memory (LTM) (called P1 rules), to managing the size of digit chains (P2 and P3 rules, which create a frame of two or three slots) and to filling these slots (if there are any empty slots, P4 rules will fill them with 0s). Separators, such as thousands and hundreds, are used to identify the number of slots; once every segment is placed in its digit form in the chain, it is transcribed. The model accounts for the development of transcoding processes through practice: experience leads to an expansion of the numerical lexicon and improvement of conversion rules.

The ADAPT model is the only cognitive model of number transcoding which makes testable predictions regarding both working memory capacity and phonological/lexical representations and their respective roles in the typical and atypical development of transcoding abilities. Moreover, even though it is not explicitly stated in the original publication (Barrouillet et al., 2004), ADAPT clearly emphasizes the importance of

phonological encoding in the first steps of number writing pro-duction, and this has not been investigated in more detail. Because both working memory and the ability to form lexical representations of numbers and - as we assume here - phone-mic awareness are related to mathematical performance, ADAPT is the only transcoding model directly examined in the present study.

Short-term memory and working memory (thereafter WM) are involved in the temporary storage of verbal information, lex-ical retrieval, and the execution of the manipulations to generate the Arabic output. Working memory representations are also involved in creating a sequence of digits and possibly blank spaces to be filled with subsequent procedures. It has been consistently related to number transcoding performance and error patterns (Camos, 2008; Zuber et al., 2009; Pixner et al., 2011). The role of working memory in transcoding tasks can be outlined in the following steps: encoding the number to be transcoded; monitor-ing the application of transcoding rules and the production of the numeral (Lochy and Censabella, 2005).

Another cognitive mechanism that may be involved in num-ber transcoding is phonemic awareness. Phonemic awareness is the subcomponent of phonological processing which is related to the ability to perceive and manipulate the phonemes that con-stitute words (Wagner and Torgesen, 1987). According to the ADAPT model (Barrouillet et al., 2004), the phonological encod-ing of the verbal numerals is the primary step in transcoding procedures, before the use of algorithm rules and retrieval from LTM. Therefore, limitations in phonological processing capacity may constrain the ability to transcode, particularly in the case of longer and more complex numbers. Phonological processing may also interact with the capacity of verbal working memory. The more demanding the phonological processing of numerical stim-uli, the fewer resources would remain available in verbal working memory for transcoding. Although the conversion of a verbal representation to an Arabic one is related to phonological repre-sentations, this association has not yet been investigated in detail in the ADAPT model.

Krajewski and Schneider (2009) found that phonological awareness facilitates the differentiation and manipulation of sin-gle words in the number word sequence. These authors built a model of early arithmetic development that postulates three dif-ferent levels: (1) basic numerical skills, in which children are already able to discriminate between quantities and to recite number words, without accessing their quantitative semantic meaning; (2) quantity-number concept, when there is a linkage between magnitudes and the number words that represent them; (3) number relationships, the point at which children understand that the difference between two numbers is another number. According to these authors, phonological awareness (measured by phoneme synthesis and rhyming tasks) plays an important role in the first level. The authors claim that because this phonological skill is related to the ability to differentiate and manipulate mean-ingful segments of language, it is also important in differentiating number words ("one," "two," "three" instead of "onetwothree").

In view of the above, the aim of this study is to investi-gate the role of specific cognitive mechanisms underlying num-ber transcoding such as general cognitive ability, verbal and non-verbal short-term and working memory, magnitude repre-sentation, and phonemic awareness. More specifically, our main goal was to investigate the relative impact of phonemic aware-ness on number transcoding. Phonemic awareness is related to reading and spelling skills (Wagner and Torgesen, 1987; Castles and Coltheart, 2004; Hulme et al., 2012; Melby-Lervå et al., 2012), and recent studies have also focused on its association with arithmetic fact retrieval and with arithmetic word problems (Hecht et al., 2001; Boets and De Smedt, 2010; De Smedt et al., 2010). Importantly, many measures of phonemic awareness, such as the phoneme elision task employed in the present investiga-tion, require a certain availability of working memory resources. Working memory is recruited in such tasks when the participant must hold a word in mind while determining the phonological information to be deleted (De Smedt et al., 2010). Both verbal and visuospatial working memory play important roles in numer-ical transcoding according to the ADAPT model (Camos, 2008; Zuber et al., 2009), but no study so far has investigated the spe-cific contribution of phonemic awareness and working memory in number transcoding tasks.

Two main hypotheses will be addressed in the present study: First, based on the central role assigned by the ADAPT model to working memory capacity (Barrouillet et al., 2004; Camos, 2008), one can argue that working memory contributes to num-ber transcoding independently because working memory capacity is putatively implicated in the use of transformation rules and procedures employed during transcoding. Second, at least part of the influence of working memory on number transcoding should be mediated by phonemic awareness. Phonemic aware-ness scores are assumed to index the quality of the underlying phonological representations. These representations are related to the perception and manipulation of sound-based processes (Simmons and Singleton, 2008); therefore, phonemic awareness performance would have an impact on verbal working memory and transcoding skills.

MATERIALS AND METHODS

The study was approved by the local research ethics committee (COEP–UFMG) and is in line with the Declaration of Helsinki. Children participated only after informed consent was obtained. Informed consent was obtained in written form from parents and orally from children.

SAMPLE

A total of 487 children in grades 2–4 were invited from public schools in Belo Horizonte, Brazil. Of these children, 207 (42%) children agreed to take part in this study. Testing was conducted in the children's own schools. The various tasks were presented in four different pseudo-random orders during one session that lasted approximately 1 h.

We excluded five children from the sample due to low intel-ligence (performance on Raven's Colored Progressive Matrices below one standard deviation). One child did not complete the entire battery and was also excluded from the analysis. Twenty-nine children were excluded from further analyses because either they had a poor R^2 on the fitting procedure to calculate their internal Weber fraction on the non-symbolic comparison task

($R^2 < 0.2$) or they showed an internal Weber fraction that exceeded the limit of discriminability of the non-symbolic magnitude comparison task ($w > 0.6$). The final sample comprised 172 children (55.2% girls), with a mean age of 111.84 months ($SD = 10.90$), ranging from 94 to 140 months.

INSTRUMENTS

The following instruments were used in the cognitive assessment: Raven's Colored Progressive Matrices, Digit Span, Corsi Blocks, Non-symbolic magnitude comparison task, Phoneme Elision and Arabic number writing task.

(a) Raven's Colored Progressive Matrices: general intelligence was assessed with the age-appropriate Brazilian validated version of Raven's Colored Matrices (Angelini et al., 1999). The analyses were based on z-scores calculated from the manual's norms.

(b) Digit Span: Verbal short-term and working memory were assessed with the Brazilian WISC-III Digit Span subtest (Figueiredo, 2002). Performance in the forward order was considered a measure of verbal short-term memory, and the backward order was used to assess verbal working memory (Figueiredo and Nascimento, 2007). We evaluated the total score (correct trials x span) in both the forward and backward orders.

(c) Corsi Blocks: This test is a measure of the visuospatial component of short-term and working memory. It consists of a set of nine blocks, which the examiner taps in a certain sequence. The test starts with sequences of two blocks and can reach a maximum of nine blocks. We used the forward and backward orders according to Kessels et al. (2000). In the forward condition, the child is instructed to tap the blocks in the same order as the examiner, and in the backward condition, in the reverse order. We also evaluated the total scores.

(d) Non-symbolic magnitude comparison task: In this task, the participants were instructed to compare two simultaneously presented sets of dots, indicating which one contained the larger number. Black dots were presented on a white circle over a black background. In each trial, one of the two white circles contained 32 dots (reference numerosity) and the other contained 20, 23, 26, 29, 35, 38, 41, or 44 dots. Each magnitude of dot sets was presented eight times. The task comprised 8 learning trials and 64 experimental trials. Perceptual variables were varied such that in half of the trials the individual dot size was held constant, while in the other half, the size of the area occupied by the dots was held constant (see exact procedure descriptions in Dehaene et al., 2005). Maximum stimulus presentation time was 4.000 ms, and the inter-trial interval was 700 ms. Before each trial, a fixation point appeared on the screen: a cross, printed in white, with each line 30 mm long. If the child judged that the right circle presented more dots, a predefined key localized in the right side of the keyboard should be pressed with the right hand. However, if the child judged that the left circle contained more dots, then a predefined key on the left side had to be pressed with the left hand (Costa et al., 2011). As a measure of the number sense acuity, the internal Weber fraction

(w) was calculated for each child based on the Log-Gaussian model of number representation (Dehaene, 2007), with the methods described by Piazza et al. (2004).

(e) Phoneme Elision: This is a widely accepted measure of phonemic awareness (Wagner and Torgesen, 1987; Castles and Coltheart, 2004; Hulme et al., 2012; Melby-Lervå et al., 2012). The child hears a word and must say what the word would be if a specified phoneme in the word were to be deleted (e.g., "*filha*" without /f/ is "*ilha*" [in English, it would be similar to "cup" without /k/ is "up"). The test comprises 28 items: in 8 items, the child must delete a vowel, and in the other 20, a consonant. The consonants to be suppressed varied by place and manner of articulation. The phoneme to be suppressed could be in different positions within the words, which ranged from 2 to 3 syllables. The internal consistency of the task is 0.92 (KR-20 formula).

(f) Arabic number writing task: To evaluate number transcoding, children were instructed to write the Arabic forms of dictated numbers. This task consists of 40 items, up to 4 digits (3 one-digit numbers, 9 two-digit numbers, 10 three-digit numbers and 18 four-digit numbers). The one- and two-digit numbers were classified as "lexical items" (12 items), and the other 28 items require the use of algorithm-based rules in order to be written (Barrouillet et al., 2004; Camos, 2008). This task has been used in a previous study with a comparable sample, and the consistency of this task was KR-20 = 0.96 (Moura et al., 2013).

ANALYSIS

The differential impact of phonemic awareness and working memory on number transcoding was investigated in a hierarchical regression analysis with Arabic number writing as the dependent variable. Age and intelligence were entered first, and working memory and the Weber fraction in a second step, using the stepwise method. The phoneme elision task was entered in the model in a third step, also using the stepwise method. This allowed us to investigate the specific contribution of phonemic awareness to number transcoding performance after working memory variance was taken into account.

As a complement, path analyses, including all measures of age, intelligence, working memory and phonemic awareness were calculated, to determine the specific contribution of phonemic awareness as a mediator of the effect of working memory on number transcoding.

RESULTS

Thirty-three percent of the children did not commit any errors in the number transcoding task. Ninety-three percent of the children did not commit any errors on the numbers that can be lexically retrieved (items 1–12). According to what is suggested by the ADAPT model, errors rates increased with the number of rules required for number transcoding. In the numbers that required 3 transcoding rules, 50% of the children committed errors, in the 4-rules, 71.6% presented some errors, in the 5-rules, 73.3% and, finally in the more complex items (6 and 7 rules), 84.5% of the children committed, at least, one error.

Since one-third of the sample did not commit any error in the transcoding task, one may argue that they should be excluded from the sample to avoid biases in the estimation of the covariance matrix, particularly with regard to the association between transcoding performance and other cognitive functions. To investigate the occurrence of bias, regression and path analyses were performed in the full sample and in the sample without the children with perfect score in the transcoding task. Results were numerically comparable in both regression and path analyses and their interpretation was exactly the same. For this reason, we decided to report the results obtained by analysing the full sample.

ASSOCIATION BETWEEN COGNITIVE VARIABLES AND TRANSCODING ABILITY

First, the specific impact of the different cognitive mechanisms on number transcoding was evaluated by means of hierarchical regression models. To approximate a normal distribution, error rates of the Arabic number writing task were arcsine transformed. Initially, we examined the general association between these measures through Pearson's correlations. Inspection of **Table 1** reveals that the error rates observed in the number transcoding task were negatively correlated to age, intelligence, working memory, and phonemic awareness. There was also a weak positive correlation between error rates in number transcoding and the Weber fraction, which may reflect the maturation level of more general numerical skills. Moreover, phonemic awareness was significantly correlated to intelligence and working memory.

To investigate in more detail the specific impact of phonemic awareness on transcoding abilities, a hierarchical regression model was calculated (**Table 2**). In this model, more general determinants of cognitive development were entered first, and more specific predictors of transcoding ability were included later on, in a hierarchical fashion. In step 1, age and intelligence were included as general factors that predict school achievement, using the enter method. In step 2, the following cognitive measures were included: Weber fraction and the total scores of the forward and backward orders of Digit Span and Corsi Blocks. Last, in step 3, we included the phoneme elision score. The stepwise method was used in steps 2 and 3 to avoid redundancy and to guarantee a high degree of parsimony.

The regression model reveals that after removing the effects of age and intelligence in step 1, verbal working memory remains a significant predictor of transcoding performance in step 2. Nevertheless, the addition of phonemic awareness to the model in step 3 leads to the exclusion of verbal working memory. Phonemic awareness, along with age and intelligence, was a significant predictor of number transcoding and absorbed the impact of verbal working memory on transcoding performance. The model explains a moderate amount of variance (**Table 2**). Measures of the approximate number system, visuospatial short-term memory, and visuospatial working memory were not retained in the model.

The reason to employ a hierarchical regression model in this analysis is to demonstrate the validity of the present experimental setup. By entering the measures of working memory in the regression model first we are able to replicate previous studies and thereby show that our measures of working memory were well-chosen and are associated to transcoding abilities. After completing this step of validation of well-established results, we continue the investigation showing that phonemic awareness absorbs the impact of measures of working memory on transcoding capacity. We have also calculated a regression model allowing the effect of phonemic awareness to vary simultaneously to measures of working memory, that is, with no hierarchical distinction

Table 2 | Regression analysis for number transcoding (errors arcsine, adjusted $r^2 = 0.41$).

Predictor	Beta	Partial t	Sig	r^2 change
Intercept		10.14	<0.001	
Age (months)	−0.404	−6.487	<0.001	0.305
Raven	−0.225	−3.282	0.001	
Digit span—backward	−0.089	−1.358	0.176	Excluded
Weber fraction	0.095	1.545	0.124	Excluded
Digit span—forward	−0.056	−0.885	0.378	Excluded
Corsi blocks—backward	−0.035	−0.529	0.598	Excluded
Corsi blocks—forward	−0.003	−0.051	0.959	Excluded
Phoneme elision	−0.337	−5.038	<0.001	0.088

Table 1 | Correlation coefficients.

	1	2	3	4	5	6	7	8
1. Age (in months)	1							
2. Raven	−0.23**	1						
3. Digit span-Forward	0.19*	0.19*	1					
4. Digit span-Backward	0.05	0.34**	0.18*	1				
5. Corsi blocks Forward	0.19*	0.28**	0.15*	0.20**	1			
6. Corsi blocks Backward	0.01	0.34**	0.14	0.36**	0.36**	1		
7. Weber fraction	−0.19*	−0.11	−0.17*	−0.19*	−0.16*	−0.13	1	
8. Phoneme elision	0.11	0.36**	0.23**	0.36**	0.24**	0.25**	−0.13	1
9. Number transcoding	−0.11	−0.17*	−0.11	−0.15*	−0.10	−0.13	0.21**	−0.36**

**Correlation is significant at the 0.01 level (2-tailed).

*Correlation is significant at the 0.05 level (2-tailed).

between these variables. Results were largely comparable with those reported previously: only phonemic awareness is retained in the model along with intelligence and age ($R^2 = 0.64$; adjusted $R^2 = 0.40$; $b = -0.02$).

DESCRIBING THE ROLES OF PHONEMIC AWARENESS AND VERBAL MEMORY IN ARABIC NUMBER TRANSCODING

As shown in the previous section, the influence of the verbal working memory on number transcoding is shared with phonemic awareness. Therefore, as a complement to the previous findings, path analyses including both working memory and phonemic awareness, as well as Weber fraction, were calculated in order to investigate the interplay of these variables in number transcoding.

To determine the strength of the effect of phonemic awareness on number transcoding, a sequence of models was calculated and compared. Chi-square and the approximate fit indexes root mean square residual (RMR), goodness of fit index (GFI), adjusted goodness of fit index (AGFI), comparative fit index (CFI) and root mean square error of approximation (RMSEA) were used to evaluate model quality. A non-significant chi-square indicates no significant discrepancy between model and data. The RMR measures the ratio of residuals in comparison to the covariances expressed by the models. Values smaller than 0.10 are considered adequate. GFI, AGFI, and CFI evaluate the degree of misspecification present in the model. Usually, the best acceptable values are greater than 0.90. Finally, the Root Mean Square Error of Approximation, or RMSEA, considers the model complexity when evaluating the model fit. The RMSEA is considered acceptable when it is lower than 0.05. The Chi-square difference between models was employed to compare models with increasing numbers of free parameters. Models were calculated in the software AMOS v.19 using the maximum likelihood estimation function.

To control for the influence of developmental and intellectual levels on the path models, we calculated the unstandardized residuals of the independent variables (short-term and working memory, Weber fraction and phonemic awareness), in which the portion of variance due to age (in months) and/or intelligence was removed. These adjusted values of working memory, magnitude processing and phonemic awareness were entered as the exogenous variables in the path analyses. All the covariances between the exogenous variables were set as free (**Figure 1**).

Those variables with negative standardized values indicate that higher scores in these predictors lead to lower error rates in the number transcoding task. The only exception is the Weber Fraction path, in which higher values indicate poorer magnitude representation acuity and, hence, more errors in number transcoding.

Fit statistics of path models are shown in **Table 3**. The first and most complex model (ALL PATHS) included the two measures of short-term and working memory (forward and backward versions of Digit Span and Corsi Blocks), as well as Weber fraction and an additional Phoneme Elision mediation path between both the forward and backward versions of the Digit Span and the number writing tasks. This model presented adequate fit indexes but is not parsimonious. Models with fewer parameters to be

estimated were designed and were compared to the ALL PATHS model and to one another.

First, the NO VISUOSPATIAL model removed the paths from visuospatial memory to transcoding. Accordingly, the NO ANS model also suppressed the path from the Weber fraction to transcoding. In one further step, two models were calculated. In the first (MEDIATION PATH), the contribution of verbal working memory to transcoding is partially mediated by phonemic awareness. Finally, to determine the relevance of phonemic awareness for transcoding, in the last model, the path from Phoneme elision to Number transcoding was removed, while the direct paths from verbal working memory to transcoding were retained (NO MEDIATION). If the exclusion of any of these paths leads to a statistically significant decrease in model fit, one may conclude that the specific parameters removed from the more parsimonious version of the path model contribute substantially to model fit.

Inspection of **Table 3** reveals that all models including the Phoneme Elision-mediation path reached satisfactory fit levels. Nevertheless, all models presented large residuals, as indicated by the RMR, which suggests that the variables included in the models were not sufficient to fully explain the variance in the number writing task. However, non-significant Chi-squares and the other fit measures associated with these models were largely acceptable.

Overall, the model that presented the worst fit indices was the one that excluded the Phoneme Elision-mediation path and assumed that Digit Span has a direct influence on number transcoding (NO MEDIATION). Model comparisons corroborate these results because the model NO MEDIATION presented statistically poorer fit than all other models. Its chi-square was statistically significant, and the model did not present any adequate fix indexes (**Table 3**). This finding suggests that phonemic awareness is a relevant predictor of transcoding performance, with substantial specific contribution. Moreover, comparisons among all other models only produced non-significant chi-square differences. Given the statistical equivalence of these models, one may select the model MEDIATION PATH, in which the effect of working memory on transcoding performance is partially mediated by phonemic awareness, as the most parsimonious description of the present data. Importantly, the association between verbal working memory and phonemic awareness is stronger than that between verbal short-term memory and phonemic awareness. Regression values of the model MEDIATION PATH are depicted in **Figure 1**.

DISCUSSION

The present study investigated the impact of phonological skills on a number transcoding task, and it is, to our knowledge, the first to simultaneously evaluate the relative impact of short-term and working memory, number sense and phonemic awareness on number transcoding. Our results revealed two main findings. First, we confirmed previous evidence of a verbal working memory effect on number transcoding, and, more importantly, we provided evidence of a relationship between number transcoding and phonemic awareness. Our second main finding is that the well-established relationship between verbal working memory capacity and number transcoding is mediated by phonemic
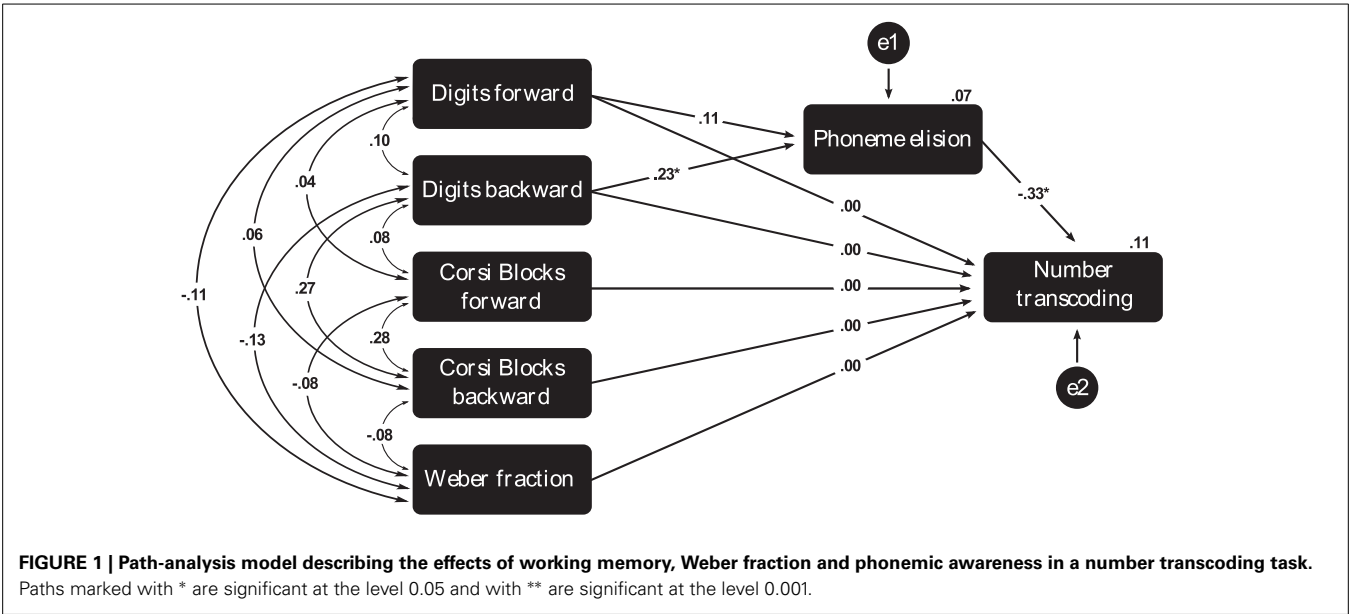
FIGURE 1 | Path-analysis model describing the effects of working memory, Weber fraction and phonemic awareness in a number transcoding task.
Paths marked with * are significant at the level 0.05 and with ** are significant at the level 0.001.

Table 3 | Fit statistics for path models regarding number writing performance.

Number transcoding model	χ2	df	p	RMR	GFI	AGFI	CFI	RMSEA
ALL PATHS	0.793	3	0.851	1.066	0.999	0.988	1.000	0.000
NO VISUOSPATIAL	0.953	5	0.966	1.067	0.998	0.991	1.000	0.000
NO ANS	2.396	6	0.880	1.067	0.996	0.981	1.000	0.000
MEDIATION PATH	4.156	8	0.843	1.071	0.993	0.976	1.000	0.000
NO MEDIATION	30.654	9	<0.001	3.388	0.954	0.858	0.563	0.119

Note: RMR, root mean square residual; GFI, goodness-of-fit index; AGFI, adjusted-goodness-of-fit index; CFI, comparative fit index; RMSEA, root mean square error of approximation.

awareness abilities. In the following sections, these topics will be discussed in more detail.

THE IMPACT OF VERBAL AND VISUOSPATIAL WORKING MEMORY ON ARABIC NUMBER WRITING

The performance of children in the number writing task was far from being flawless. They present many errors on the more complex two-, three-, and four-digit items, which require more than three transcoding rules, according to ADAPT. These findings are in accordance to what has been reported in the literature regarding transcoding skills of school aged children (Moura et al., 2013) and have been interpreted as a product of working memory processes in number transcoding (Camos, 2008). However, little is effectively known about the selective impact of different components of working memory on number transcoding. To our knowledge, this was the first study to analyze this problem in greater depth. Although a specific role of the central executive function in transcoding has been suggested (Camos, 2008), the present study is the first to explore the impact of phonological and visuospatial working memory in a number writing task and distinguish them from the central executive. We provide evidence regarding the specific role of phonological working memory and, more precisely, of the quality of underlying

phonological representations, by means of the phonemic awareness performance.

Working memory plays an important role in the algorithmic-based procedures of number transcoding (Camos, 2008; Pixner et al., 2011). Essentially, it is believed to be involved in the maintenance of verbal units from the verbal numbers and in managing the new digit chain. In our study, we found that better verbal working memory capacity was associated with higher number transcoding performance. Interestingly, the same does not apply to the visuospatial components of short-term and working memory, as none of them revealed an association with transcoding performance in correlation, regression or path analyses. In a previous study by Zuber et al. (2009), the visuospatial working memory component was associated with the management of Arabic code syntax. Nevertheless, it is important to note here that the sample used in this other study was composed of German-speaking first graders, and the German number word system is different from the Portuguese system. In German, the order of the units and decades in the verbal numerals is inverted in comparison to the Arabic ones. One possibility, therefore, is that transcoding numbers in Portuguese demands less visuospatial working memory capacity than in languages with this inversion. Linguistic comparison research remains necessary to confirm this hypothesis.

Raghubar et al. (2010) reviewed evidence indicating that the influence of the subcomponents of working memory on arithmetic performance might vary according to age. The visuospatial component is recruited in earlier phases of development, while children are still learning basic mathematical concepts, whereas the phonological loop is more relevant after these skills have already been mastered. Although Raghubar et al. (2010) did not specifically discuss number transcoding, this study reviews evidence regarding the complex and dynamic nature of the relationship between working memory and math achievement. Consistent with these results, no effect of visuospatial working memory on number transcoding was observed in second- to fourth-grade children in the present study.

THE RELATIONSHIP BETWEEN VERBAL WORKING MEMORY AND PHONEMIC AWARENESS

The first step of writing Arabic numbers from dictation proposed by the ADAPT model (Barrouillet et al., 2004) is the phonological encoding of the auditory input, which consists of verbal numerals. Nevertheless, the procedures involved in this phonological encoding are still not completely specified. Here we showed that, in addition to working memory capacity, phonemic awareness also plays an important role in number transcoding. Our results showed that even when considering the influence of working memory and basic numerical skills on number transcoding, the predictive value of phonemic awareness abilities was substantial. This suggests that phonemic awareness is an important facilitator of the phonological encoding required in the initial steps of number transcoding.

Another aim of the present study was to clarify the influence of phonemic awareness on number transcoding. We aimed to investigate whether there is a direct influence of verbal working memory on number transcoding or if this association would be mediated by phonemic awareness. Our results presented evidence showing that phonemic awareness mediates the influence of verbal working memory in number transcoding, even after controlling for the effects of age and intelligence. In the path analyses, the removal of the Phoneme Elision-mediation path had a deleterious effect on model fit, which suggests that this parameter contributes crucially to improve the model fit.

This finding is consistent with the ADAPT model, which postulates that the first step in number transcoding would be the encoding of the verbal string into its phonological form (Barrouillet et al., 2004). This encoding phase would be followed by parsing procedures that segment these strings into smaller units. Smaller units are then sequentially processed through a production system in which verbal working memory is required for transcoding algorithms. It is possible to hypothesize that phonemic awareness would be the main cognitive precursor engaged in the phonological encoding phase that precedes further verbal working memory involvement in number transcoding.

A plausible explanation for the association between phonemic awareness and the influence of verbal working memory in number transcoding is the "weak phonological representation hypothesis" (Simmons and Singleton, 2008). According to this model, phonological processing deficits would impair the quality of phonological representations and thus affect aspects of numerical cognition that involve the manipulation of a verbal code.

The performance in verbal working memory and phonemic awareness depend on the same underlying and latent phonological representations (Hecht et al., 2001; Alloway et al., 2005; Durand et al., 2005). In our study, it was also possible to observe this association through the positive correlation between verbal working memory and phonemic awareness. Baddeley et al. (1975) had already suggested that, given that verbal short-term memory is a speech-based system, its capacity should be measured in more basic speech units, such as phonemes. Oakhill and Kyle (2000) also found that phonemic awareness (operationalized by means of phoneme elision and phoneme segmentation tasks) had a strong association with word and sentence span.

Evidence indicates that the influences of phonemic awareness and verbal working memory on literacy acquisition are both shared and unique (Mann and Liberman, 1984; Alloway et al., 2005). Factor analytical studies indicate that different types of phonological awareness tasks are loaded onto a single latent construct (Schatschneider et al., 1999). Tasks vary, however, in the additional cognitive demands they impose, regarding, for instance, working memory and other general cognitive components. According to this type of reasoning, different phonemic awareness tasks assess a common phonological processing construct plus additional varying components that change according to task demands. A task such as phoneme elision would consist then of at least two components, one tapping the phonological latent construct and the other one depending on working memory demands. Previous studies (Oakhill and Kyle, 2000; Alloway et al., 2005) have investigated the influence of verbal working memory on phonemic awareness performance. This question, however, is rather complex and our results emphasize the importance of also investigating the other direction of this relationship. This is especially relevant regarding the interplay between verbal working memory, phonemic awareness and number transcoding skills.

Another dimension adding complexity to the relationship between phonemic awareness and verbal working memory is the child's individual level of development, which may be characterized as the degree of automatization in phonological processing. Before the child acquires expertise with phonemic awareness, a task such as phoneme elision may impose heavy demands on the central executive. As the child progressively acquires experience with phonological processing, this task can be solved in a more automatic way, freeing working memory resources for other tasks relevant for more advanced operations. If, however, the child does not acquire abilities of accurately and automatically processing the phonemic units, precious working memory resources will be less available for numerical transcoding. Accurate and automatic phonemic processing liberates sparse processing resources necessary to solve more complex tasks.

Disclosing a complex relationship among working memory, phonemic awareness and transcoding has important consequences for math achievement in general and for its disorders. School achievement in reading and/or mathematics depends

on a complex interaction between general and specific cognitive factors. As the child acquires expertise in specific domains, such as phonemic and/or quantitative representations, processing resources are liberated to work in increasingly more complex activities. The accurate and automatic nature of more basic sound and quantitative representations may thus influence the whole process of school learning, explaining variances both in achievement and in working memory. Johnson (2012) recently proposed that the occurrence of learning disabilities depends on such an interaction between specific and general cognitive factors. If a specific impairment, say in phonological or number processing, can be compensated by central executive resources, there is a smaller probability that the individual develops a learning disability. Otherwise, if executive processing resources are not sufficient to compensate or automatize basic cognitive processes, difficulties persist. This hypothesis has been explored in another report, investigating two cases of math learning difficulties (Haase et al., in press, this issue). In one case, math learning difficulties were associated with a lack of automatization and in the other case with impaired executive working memory resources.

There have been few studies that directly addressed the relationship between verbal memory and phonemic awareness during the performance of arithmetic tasks. Leather and Henry (1994) claim that both constructs share a certain amount of variance with arithmetic performance because phonemic manipulation demands arithmetical processes (for instance, phoneme elision tasks require, literally, the subtraction of a sound) and also involve working memory for the mental retention and management of verbal information. Phoneme elision tasks require both storage and processing of phoneme units because children usually hold the word in mind while deleting one sound and producing the new word with what is left (Oakhill and Kyle, 2000). Hecht et al. (2001) longitudinally investigated the role of phonological awareness in arithmetic development of children from different age ranges and found that from the 3rd to 4th grades, as well as from the 4th to 5th grades, this was the only subcomponent of phonological processing that explained the growth of performance in a standardized arithmetic task. According to the authors, the same memory resources engaged in arithmetic problem solving are also recruited in phonological awareness tasks.

Our findings are in accordance to what was reported by Michalczyk et al. (2013). The authors also found that the simultaneous inclusion of verbal and visuospatial working memory, the central executive as well as phonological awareness in a regression model showed that only phonological awareness—none of the working memory subcapacities—had a direct impact on basic quantity-number competencies. In this study, they investigated the performance of children aged 5 and 6 in a number sequence task, in which children had to recite the number word sequence forwards up to 31 and backwards from 5. Afterwards they had to name 3 subsequent and 3 preceding number words. Even though they did not use a transcoding task, one can infer from this result that phonological awareness might mediate the relation between verbal working memory and number words knowledge. Nevertheless, as mentioned above, our study was the first one to provide evidence regarding the mediation of the effect of

verbal working memory on number transcoding by phonemic awareness.

FINAL REMARKS

Mathematics encompasses a range of several different competences, such as numerical estimation, word problems, fact retrieval and number transcoding. Standardized arithmetic tasks usually assess these different abilities simultaneously and do not tap their specificities. It is important to investigate the distinct cognitive mechanisms that are associated with each of these mathematical skills. In our study, we concluded that phonemic awareness and verbal memory are directly connected to number transcoding, being important pathways between the verbal input and the transcription of the Arabic output.

The acuity of number sense, as measured by the Weber Fraction, did not influence number writing, suggesting that the assessment of numerical magnitude is not a necessary step in number transcoding. The acuity of number sense has been considered an important predictor of arithmetic performance (Halberda et al., 2008), but its relationship to number transcoding is less explored.

Although we did not explicitly assess children with learning disabilities, our results provide additional support to the hypothesis that phonemic awareness might be a cognitive mechanism that underlies both dyslexia and dyscalculia. Epidemiological studies describe high comorbidity rates between reading and mathematical difficulties: approximately 40% of dyslexics also have arithmetic difficulties (Lewis et al., 1994), and the prevalence of dyslexia and dyscalculia is similar, approximately 4–7% (Dirks et al., 2008; Landerl and Moll, 2010). The finding that phonemic awareness is related to number transcoding is useful in the comprehension of mathematical difficulties presented by dyslexic children (Haase et al., in press, this issue). We suggest that this should also be assessed in neuropsychological evaluations as well as in clinical interventions for children with learning disabilities.

ACKNOWLEDGMENTS

REFERENCES

Alloway, T. P., Gathercole, S. E., Adams, A., Willis, C., Eaglen, R., and Lamont, E. (2005). Working memory and phonological awareness as predictors of progress towards early learning goals at school entry. *Br. J. Dev. Psychol.* 23, 417–426. doi: 10.1348/026151005X26804

Angelini, A. L., Alves, I. C. B., Custódio, E. M., Duarte, W. F., and Duarte, J. L. M. (1999). *Matrizes Progressivas Coloridas de Raven: Escala Especial. Manual.* São Paulo: CETEPP.

Baddeley, A. D., Thompson, N., and Buchanan, M. (1975). Word length and the structure of short-term memory. *J. Verb. Learn. Verb. Behav.* 14, 575–589. doi: 10.1016/S0022-5371(75)80045-4

Barrouillet, P., Camos, V., Perruchet, P., and Seron, X. (2004). ADAPT: a developmental, asemantic, and procedural model for transcoding from verbal to arabic numerals. *Psychol. Rev.* 111, 368–394. doi: 10.1037/0033-295X.111.2.368

Boets, B., and De Smedt, B. (2010). Single-digit arithmetic in children with dyslexia. *Dyslexia* 16, 183–191. doi: 10.1002/dys.403

Camos, V. (2008). Low working memory capacity impedes both efficiency and learning of number transcoding in children. *J. Exp. Child Psychol.* 99, 37–57. doi: 10.1016/j.jecp.2007.06.006

Castles, A., and Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition* 91, 77–111. doi: 10.1016/S0010-0277(03)00164-1

Costa, A. J., Silva, J. B. L., Chagas, P. P., Krinzinger, H., Lonneman, J., Willmes, K., et al. (2011). A hand full of numbers: a role for offloading in arithmetics learning? *Front. Psychol.* 2:368. doi: 10.3389/fpsyg.2011.00368

Dehaene, S., Izard, I., and Piazza, M. (2005). *Control Over Non-Numerical Parameters in Numerosity Experiments*. Available online at: www.unicog.org/docs/DocumentationDotsGeneration.doc

Dehaene, S. (2007). "Symbols and quantities in parietal cortex: elements of a mathematical theory of number representation and manipulation," in *Sensorimotor Foundations of Higher Cognition – Attention and Performance XXII*, eds P. Haggard, Y. Rossetti, and M. Kawato (Cambridge, MA: Harvard University Press), 527–574.

Deloche, G., and Seron, X. (1987). "Numerical transcoding: a general production model," in *Mathematical Disabilities. a Cognitive Neuropsychological Perspective*, eds G. Deloche and C. Seron (Orgs.).(Hillsdale, NJ: Erlbaum), 137–170.

De Smedt, B., Taylor, J., Archibald, L., and Ansari, D. (2010). How is phonological processing related to individual differences in children's arithmetic skills? *Dev. Sci.* 13, 508–520. doi: 10.1111/j.1467-7687.2009.00897.x

Dirks, E., Spyer, G., van Lieshout, E. C., and de Sonneville, L. (2008). Prevalence of combined reading and arithmetic disabilities. *J. Learn. Disabil.* 41, 460–473. doi: 10.1177/0022219408321128

Durand, M., Hulme, C., Larkin, R., and Snowling, M. (2005). The cognitive foundations of reading and arithmetic1 skills in 7- to 10-year-olds. *J. Exp. Child Psychol.* 91, 137–157. doi: 10.1016/j.jecp.2005.01.003

Fayol, M., and Seron, X. (2005). "About numerical representations," in *Handbook of Mathematical Cognition*, eds J. I. D. Campbell (New York, NY: Psychology Press), 3–22.

Figueiredo, V. L. M. (2002). *WISC-III: Escala de Inteligência Wechsler para Crianças. Manual Adaptação e Padronização Brasileira*. São Paulo: Casa do Psicólogo.

Figueiredo, V. L. M., and Nascimento, E. (2007). Desempenhos nas duas tarefas do subteste dígitos do WISC-III e do WAIS-III. *Psicol. Teor. Pes.* 23, 313–318. doi: 10.1590/S0102-37722007000300010

Geary, D. C. (2000). From infancy to adulthood: the development of numerical abilities. *Eur. Child Adolesc. Psychiatry* 9, 11–16. doi: 10.1007/s007870070004

Haase, V. G., Júlio-Costa, A., Lopes-Silva, J. B., Starling-Alves, I., Antunes, A. M., Pinheiro-Chagas, P., et al. (in press). Contributions from specific and general factors to unique deficits: two cases of mathematics learning difficulties. *Front. Psychol.*

Halberda, J., Mazzocco, M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* 455, 665–668. doi: 10.1038/nature07246

Hecht, S. A., Torgesen, J. K., Wagner, R. K., and Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: a longitudinal study from second to fifth grades. *J. Exp. Child Psychol.* 79, 192–227 doi: 10.1006/jecp.2000.2586

Hulme, C., Bowyer-Crane, C., Carroll, J., Duff, F., and Snowling, M. (2012). The causal role of phoneme awareness and letter-sound knowledge in learning to read. *Psychol. Sci.* 23, 572–577. doi: 10.1177/0956797611435921

Johnson, M. H. (2012). Executive function and developmental disorders: the flip side of the coin. *Trends Cogn. Sci.* 16:9, 454–457. doi: 10.1016/j.tics.2012.07.001

Kessels, R. P. C., Van Zandvoort, M. J. E., Kapelle, L. J., Postma, A., and De Haan, E. H. (2000). The Corsi block-tapping task: standardization and normative data. *Appl. Neuropsychol.* 7, 252–258. doi: 10.1207/S15324826AN0704_8

Krajewski, K., and Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: findings from a 3-year longitudinal study. *J. Exp. Child Psychol.* 103, 516–531. doi: 10.1016/j.jecp.2009.03.009

Landerl, K., and Moll, K. (2010). Comorbidity of learning disorders: prevalence and familial transmission. *J. Child Psychol. Psychiatry* 51, 287–294. doi: 10.1111/j.1469-7610.2009.02164.x

Leather, C. V., and Henry, L. A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *J. Exp. Child Psychol.* 58, 88–111. doi: 10.1006/jecp.1994.1027

Lewis, C., Hitch, G. J., and Walker, P. (1994). The prevalence of specific arithmetic difficulties and specific reading difficulties in 9- to 10-year-old boys and girls. *J. Child Psychol. Psychiatry* 35, 283–292. doi: 10.1111/j.1469-7610.1994.tb01162.x

Lochy, A., and Censabella, S. (2005). "Le système simbolique arabique: acquisition, évaluation, et pistes rééducatives," in *Marie-Pascale Noël*, eds *La Dyscalculie* (Marseille: Solal) 77–104.

Mann, V. A., and Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *J. Learn. Disabil.* 17, 592–598. doi: 10.1177/002221948401701005

Melby-Lervå, M., Lyster, S. H., and Hulme, C. (2012). Phonological skills and their role in learning to read: a meta-analytic review. *Psychol. Bull.* 138, 322–352. doi: 10.1037/a0026744

Michalczyk, K., Krajewski, K., Preßler, A. L., and Hasselhorn, M. (2013). The relationships between quantity-number competencies, working memory, and phonological awareness in 5- and 6-year-olds. *Br. J. Dev. Psychol.* 31, 408–424. doi: 10.1111/bjdp.12016

Moura, R., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., et al. (2013). Transcoding abilities in typical and atypical mathematics achievers: the role of working memory, procedural and lexical competencies. J. Exp. Child Psychol. 116, 707–727. doi: 10.1016/j.jecp.2013.07.008

Oakhill, J., and Kyle, F. (2000). The relation between phonological awareness and working memory. *J. Exp. Child Psychol.* 75, 152–164. doi: 10.1006/jecp.1999.2529

Piazza, M., Izard, V., Pinel, P., LeBihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human parietal cortex. *Neuron* 44, 547–555 doi: 10.1016/j.neuron.2004.10.014

Pixner, S., Zuber, J., Heřmanová, V., Kaufmann, L., Nuerk, H. C., and Moeller, K. (2011). One language, two number-word systems and many problems: numerical cognition in the Czech language. *Res. Dev. Disabil.* 32, 2683–9 doi: 10.1016/j.ridd.2011.06.004

Raghubar, K. P., Barnes, M. A., and Hecht, S. (2010). Working memory and mathematics: a review of developmental, individual difference, and cognitive approaches. *Learn. Individ. Differ.* 20, 110–122. doi: 10.1016/j.lindif.2009.10.005

Schatschneider, C., Francis, D. J., Foorman, B. R., Fletcher, J. M., and Mehta, J. (1999). The dimensionality of phonological awareness: an application of item response theory. *J. Educ. Psychol.* 91, 439. doi: 10.1037/0022-0663.91.3.439

Simmons, F. R., and Singleton, C. (2008). Do weak phonological representations impact on arithmetic development? a review of research into arithmetic and dyslexia. *Dyslexia* 14, 77–94. doi: 10.1002/dys.341

Wagner, R. K., and Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychol. Bull.* 101, 192–212. doi: 10.1037/0033-2909.101.2.192

Wood, G., Nuerk, H. C., Freitas, P., Freitas, G., and Willmes, K. (2006). What do semi-illiterate adults know about 2-digit arabic numbers? *Cortex* 42, 48–56. doi: 10.1016/S0010-9452(08)70321-6

Zuber, J., Pixner, S., Moeller, K., and 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. J. Exp. Child Psychol. 102, 60–77. doi: 10.1016/j.jecp.2008.04.003