NEURO-COGNITIVE ARCHITECTURE OF NUMERICAL COGNITION AND ITS DEVELOPMENT

EDITED BY: Elise Klein, Korbinian Moeller, Reuven Babai and Anja Ischebeck
PUBLISHED IN: Frontiers in Human Neuroscience and Frontiers in Psychology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88966-920-2
DOI 10.3389/978-2-88966-920-2

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: frontiersin.org/about/contact

NEURO-COGNITIVE ARCHITECTURE OF NUMERICAL COGNITION AND ITS DEVELOPMENT

Topic Editors:
Elise Klein, UMR8240 Laboratoire de Psychologie du Développement et de l'Education de l'enfant (LaPsyDÉ) Paris, France
Korbinian Moeller, Loughborough University, United Kingdom
Reuven Babai, Tel Aviv University, Israel
Anja Ischebeck, University of Graz, Austria

Citation: Klein, E., Moeller, K., Babai, R., Ischebeck, A., eds. (2021). Neuro-cognitive Architecture of Numerical Cognition and Its Development. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-920-2

Table of Contents

frontiers in Human Neuroscience

EDITORIAL
published: 28 April 2021
doi: 10.3389/fnhum.2021.670460

Editorial: Neuro-cognitive Architecture of Numerical Cognition and Its Development

Elise Klein[1,2]*, *Reuven Babai*[3], *Anja Ischebeck*[4] *and Korbinian Moeller*[2,5]

[1] Université de Paris, LaPsyDÉ, CNRS, Sorbonne Paris Cité, Paris, France, [2] Leibniz-Institut fuer Wissensmedien, Tuebingen, Germany, [3] Department of Mathematics, Science and Technology Education, Tel Aviv University, Tel Aviv, Israel, [4] Institute of Psychology, University of Graz, Graz, Austria, [5] Mathematics Education Centre, School of Science, Loughborough University, Loughborough, United Kingdom

Keywords: numerical cognition, numerical development, number processing, neural correlates, neuro-cognitive plasticity, developmental dyscalculia, numerical training

Editorial on the Research Topic

Neuro-cognitive Architecture of Numerical Cognition and Its Development

OPEN ACCESS

Edited and reviewed by:
Lutz Jäncke,
University of Zurich, Switzerland

Correspondence:
Elise Klein
elise.klein@u-paris.fr

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *21 February 2021*
Accepted: *26 March 2021*
Published: *28 April 2021*

Citation:
Klein E, Babai R, Ischebeck A and Moeller K (2021) Editorial: Neuro-cognitive Architecture of Numerical Cognition and Its Development. Front. Hum. Neurosci. 15:670460. doi: 10.3389/fnhum.2021.670460

Living in our modern societies requires being numerically literate not only in everyday life (e.g., reading the clock, dealing with money, etc.) but also in educational (e.g., mathematics and science classes) and professional contexts (e.g., accounting but also crafts). Hence, numerical and mathematical skills were repeatedly observed to predict not only occupational success (e.g., Ritchie and Bates, 2013) but also more general life prospects (Parsons and Bynner, 2005). However, insufficient numerical and mathematical skills still affect a considerable share of students—even in developed Western societies as reflected in international comparison studies such as PISA (OECD, 2016).

In recent years, numerical cognition research made considerable progress in describing and understanding cognitive processes and succeeded in specifying the neural architecture underlying numerical cognition. On the basis of these developments, we invited empirical and theoretical contributions for a Research Topic on the *Neuro-cognitive Architecture of Numerical Cognition and Its Development*. We are grateful to all authors for their high-quality contributions, the reviewers for their constructive feedback and helpful suggestions during the interactive peer-review, and the publisher's editorial team for their excellent support.

The 15 contributions to our Research Topic from internationally leading groups cover different aspects of numerical cognition in children, adolescents and adults, as well as its theoretical background. The applied methods run the gamut from behavioral to neuroimaging studies, from cross-sectional to longitudinal and intervention studies. The different contributions nicely illustrate that *numerical cognition* is not a unitary and closely circumscribed construct, comprehensively accounting for the evidence observed. Instead, the empirical and theoretical contributions indicate, that there are domain-specific aspects of numerical cognition (e.g., number sense, arithmetic, spatial-numerical associations, etc.) that are influenced by domain-general abilities or traits (e.g., visuospatial abilities, self-regulation, and math anxiety).

A first set of studies investigated domain-specific and domain-general predictors of later numerical or mathematical skills. In a longitudinal behavioral study, Finke et al. observed that early non-symbolic numerical skills predicted later arithmetic skills by facilitating the acquisition of symbolic number processing. The authors concluded that non-symbolic numerical skills are foundational skills to mathematical development. Braeuning et al. proposed a multifactorial structure of early numeracy, which they tested using a confirmatory factor analysis approach on longitudinal large-scale assessment data. The authors identified four specific basic numerical skills (i.e., number sense, arithmetic, patterning/geometry, and data analysis/statistics) that remained stable from 5 to 6 years of age and thus seem to serve as building blocks for further numerical development. Schild et al. evaluated the causal influence of finger-based numerical training on initial arithmetic skills in 5 year-old children. They did not observe an advantage for the finger-based training, which they attributed to domain-general processes such as sequencing that were also required in the control training. In a cross-cultural behavioral study, Nemati et al. found an association between domain-general personality traits such as self-regulation and mathematical performance that was similar for German and Iranian students. Finally, in a commentary on the idea of a mental number line in human newborns as proposed by Di Giorgio et al. (2019), Felisatti et al. postulated the Brain's Asymmetric Frequency Tuning (BAFT) hypothesis. The BAFT hypothesis emphasizes the relevance of spatial frequencies (SFs) for spatial numerical associations and embodied numerical representations in humans. The authors posited that spatial-numerical associations in newborns (1) may be driven by absolute and relative spatial frequency processing, (2) generalize across cultures and species, and (3) may be different in newborns predisposed to atypical numerical development.

The relationship between brain structure and numerical processing abilities was investigated by two contributions: McCaskey et al. used voxel-based morphometry in a longitudinal study comparing children with developmental dyscalculia with typically developing children. The authors observed that dyscalculic children between 8 and 11 years of age had persistently reduced gray and white matter densities in brain areas associated with number processing. Heidekum et al. employed surface-based morphometry in a large sample of adults. They found associations of cortical surface properties such as sulcal depth with numerical intelligence, complex arithmetic ability, and higher-order mathematical knowledge.

Two studies investigated more applied aspects of (numerical) learning using tablet-based applications. Kohn et al. evaluated their adaptive training program Calcularis 2.0, which can be used to support dyscalculic children. The training yielded improved arithmetic and spatial number processing skills after only 12 weeks, which were still observable 3 months post-training. Jung, Meinhardt et al. evaluated a tablet-based assessment of early visuospatial abilities (VSA) using the application (app) MaGrid®. Their results substantiated the hypothesized factor structure of VSA proposed in the taxonomy of Newcombe and Shipley (2015) and provide evidence for a hierarchical development of VSA as assessed using MaGrid®.

A comparatively larger set of contributions investigated the foundations of mental arithmetic. The first three studies addressed the distinction between number magnitude manipulation (e.g., involved in calculation or number magnitude comparison) and arithmetic fact retrieval (e.g., as in simple multiplication). In a brain stimulation study employing rTMS, Fresnoza et al. investigated the role of the horizontal intraparietal sulcus (assumed to be involved in magnitude processing) and the left angular gyrus (assumed to be involved in arithmetic fact retrieval). The authors found that the involvement of these brain areas seemed to be modulated by the solution strategy employed (i.e., retrieval vs. calculation) rather than by the arithmetic operation *per se* (i.e., subtraction vs. multiplication). This means that solving less well-established multiplication problems was rather associated with the magnitude network, while solving highly overlearned subtraction problems was associated with the retrieval network. This interpretation fits well with the behavioral results of Jung, Moeller et al. Using a hemifield paradigm, the authors investigated a possible left lateralization of the representation of arithmetic facts in a number bisection task. The authors did not observe such a lateralization. They hypothesized that this might be due to the complexity of the task, so that participants might not have relied solely on arithmetic fact retrieval. However, the third study in this collection, by Suárez-Pellicioni et al. seems to contradict the notion that simple subtraction problems are solved by retrieval. The authors found in a longitudinal study using functional magnetic resonance imaging (fMRI) that children do not switch to retrieval when subtractions become more and more overlearned. Instead, it seemed that procedures became more automatic with skill development.

Another subset of studies investigated the neuro-cognitive correlates of arithmetic learning. Mosbacher et al. employed frontal and parietal anodal transcranial direct current stimulation (tDCS). While stimulation did not lead to an overall improvement of arithmetic performance, the authors observed that only frontal stimulation accelerated training gains. Additionally, Belkacem et al. compared abacus experts and non-experts with regard to their arithmetic performance and EEG activity. While the groups shared some neural signatures, they found differences indicating that abacus experts developed a new computational pathway by assigning number representations onto an imaginative abacus representation, which involved parallel activation of calculation-related areas.

Finally, in a theoretical contribution, Testolin elaborated on how computer simulation at cross-disciplinary intersections can help understand how numerical concepts are learned by the human brain. Although deep learning models are not yet designed to simulate higher-level mathematical thinking, they have the potential to grasp the acquisition of numerical concepts in its full complexity.

The broad range of studies presented in this collection documents the significant progress made in understanding different aspects of numerical cognition and development. The various methods used are state of the art and testify to the expertise of the researchers involved. The current Research Topic brought together expertise of researchers from different backgrounds which clearly advanced our understanding of numerical cognition and development and has the potential to contribute to processes of learning and teaching numerical and mathematical skills.

AUTHOR CONTRIBUTIONS

EK and KM drafted the manuscript. All authors read, corrected and approved the final manuscript.

FUNDING

This research was supported by the German Research Foundation (DFG) to EK [KL 2788/2-1].

REFERENCES

Di Giorgio, E., Lunghi, M., Rugani, R., Regolin, L., Dalla Barba, B., Vallortigara, G., et al. (2019). A mental number line in human newborns. *Dev. Sci.* 22:e12801. doi: 10.1111/desc.12801

Newcombe, N. S., and Shipley, T. F. (2015). "Thinking about spatial thinking: new typology, new assessments," in *Studying Visual and Spatial Reasoning for Design Creativity* Dodrecht, J. S. G., ed. (Springer), 179–192. doi: 10.1007/978-94-017-9297-4_10

OECD (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing. doi: 10.1787/9789264266490-en

Parsons, S., and Bynner, J. (2005). *Does Numeracy Matter More?* London: National Research and Development Centre for Adult Literacy and Numeracy.

Ritchie, S. J., and Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychol. Sci.* 24, 1301–1308. doi: 10.1177/0956797612466268

Conflict of Interest: 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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 11 February 2020
doi: 10.3389/fnhum.2020.00017

Effects of Anodal tDCS on Arithmetic Performance and Electrophysiological Activity

Jochen A. Mosbacher[1]*, *Clemens Brunner*[1], *Michael A. Nitsche*[2,3] and *Roland H. Grabner*[1,4]

[1] Educational Neuroscience, Institute of Psychology, University of Graz, Graz, Austria, [2] Department of Psychology and Neurosciences, Leibniz Research Centre for Working Environment and Human Factors, Dortmund, Germany, [3] Department of Neurology, University Medical Hospital Bergmannsheil, Ruhr-University Bochum, Bochum, Germany, [4] BioTechMed-Graz, Graz, Austria

OPEN ACCESS

Edited by:
Elise Klein,
University of Tübingen, Germany

Reviewed by:
Theodore Zanto,
University of California, San Francisco, United States
Vera Moliadze,
University Medical Center Schleswig-Holstein, Germany

***Correspondence:**
Jochen A. Mosbacher
jochen.mosbacher@uni-graz.at

Specialty section:
This article was submitted to Brain Imaging and Stimulation, a section of the journal Frontiers in Human Neuroscience

Received: *17 October 2019*
Accepted: *17 January 2020*
Published: *11 February 2020*

Citation:
Mosbacher JA, Brunner C, Nitsche MA and Grabner RH (2020) Effects of Anodal tDCS on Arithmetic Performance and Electrophysiological Activity. Front. Hum. Neurosci. 14:17. doi: 10.3389/fnhum.2020.00017

Arithmetic abilities are among the most important school-taught skills and form the basis for higher mathematical competencies. At the same time, their acquisition and application can be challenging. Hence, there is broad interest in methods to improve arithmetic abilities. One promising method is transcranial direct current stimulation (tDCS). In the present study, we compared two anodal tDCS protocols in their efficacy to improve arithmetic performance and working memory. In addition, we investigated stimulation-related electrophysiological changes. Three groups of participants solved arithmetic problems (additions and subtractions) and an n-back task before, during, and after receiving either frontal or parietal anodal tDCS (25 min; 1 mA) or sham stimulation. EEG was simultaneously recorded to assess stimulation effects on event-related (de-) synchronisation (ERS/ERD) in theta and alpha bands. Persons receiving frontal stimulation showed an acceleration of calculation speed in large subtractions from before to during and after stimulation. However, a comparable, but delayed (apparent only after stimulation) increase was also found in the sham stimulation group, while it was absent in the group receiving parietal stimulation. In additions and small subtractions as well as the working memory task, analyses showed no effects of stimulation. Results of ERS/ERD during large subtractions indicate changes in ERS/ERD patterns over time. In the left hemisphere there was a change from theta band ERD to ERS in all three groups, whereas a similar change in the right hemisphere was restricted to the sham group. Taken together, tDCS did not lead to a general improvement of arithmetic performance. However, results indicate that frontal stimulation accelerated training gains, while parietal stimulation halted them. The absence of general performance improvements, but acceleration of training effects might be a further indicator of the advantages of using tDCS as training or learning support over tDCS as a sole performance enhancer.

Keywords: arithmetic, fact retrieval, procedural calculation, EEG, transcranial direct current stimulation

INTRODUCTION

Basic arithmetic skills belong to the most important skills for educational achievement and for everyday life in general (Neisser et al., 1996; Parsons and Bynner, 1997, 2005). Not only do they allow performing simple calculations, but they also form the basis for higher mathematical competencies and understanding (e.g., Geary, 2011; Price et al., 2013). However, learning and application of arithmetic and mathematical abilities can be challenging. This is especially true for people suffering from developmental dyscalculia, with prevalence rates of 3–6% being one of the most common learning disorders (Shalev et al., 2000). But even in the general population, 22.7% perform at proficiency level 1 or below (being only able to carry out simple mathematical processes) according to an OECD survey of adult skills (OECD., 2016). As deficits in arithmetic and mathematical skills place a large burden on the individual, the interest in means to support and improve these abilities is constantly growing (e.g., Parsons and Bynner, 2005). In this study, we investigated the effects of anodal transcranial direct current stimulation (a-tDCS), a non-invasive brain stimulation method, on arithmetic abilities and related oscillatory brain activity. Performance was assessed prior, during, and after stimulation to investigate stimulation-induced improvements, EEG was recorded concomitantly in order to investigate oscillatory correlates of arithmetic performance and stimulation-induced changes.

On the behavioral level, processing of arithmetic problems relies largely on the application of two sets of knowledge: declarative knowledge about arithmetic facts and procedural knowledge about arithmetic operations (Campbell and Xue, 2001; Grabner and De Smedt, 2011). Small, easy problems (e.g., additions with sums ≤ 10) are mostly solved by fact retrieval (direct recall of the solution, an arithmetic fact, from memory), reflecting a fast and largely effortless process. Large, more complex problems (e.g., two-digit/two-digit subtractions), in contrast, are primarily solved by the application of procedural strategies (based on knowledge of arithmetic operations), which is slower and more effortful (Campbell and Xue, 2001; Destefano and LeFevre, 2004). For instance, solving a two-digit/two-digit subtraction might involve the breakdown of the problem into smaller steps ($56–27 \rightarrow 56–20 = 36 \rightarrow 36–6 = 30 \rightarrow 30–1 = 29$). These procedural calculation processes incorporate additional, domain-general functions like working memory (WM) more strongly than fact retrieval processes (Destefano and LeFevre, 2004).

On a neurophysiological level, mental arithmetic requires the interplay of a wide network of brain sites (see Menon, 2015), whereby fact retrieval is accompanied by higher activation in the angular gyrus (AG), while procedural calculation is associated with stronger activation of a broad network, including the dorsolateral prefrontal cortex (DLPFC) and the intraparietal sulcus (IPS; Grabner et al., 2009). Previous EEG research demonstrated a clear distinction between fact retrieval and procedural calculation in oscillatory event-related (de-)synchronization (ERS/ERD) patterns. Already in an early study, higher theta band power in the left hemisphere has been associated with fact retrieval during arithmetic problem solving (Earle et al., 1996). This has later been corroborated by studies showing that the processing of small (fact retrieval) problems was accompanied by stronger left hemispheric theta band ERS, while the processing of larger (procedural) problems led to less theta band ERS, but stronger alpha band ERD, especially over bilateral parieto-occipital areas (De Smedt et al., 2009; Grabner and De Smedt, 2011). Further support for this distinction comes from a training study, showing an increase in theta band ERS and a decrease in lower alpha band ERD with increasing use of fact retrieval over procedural calculation in complex arithmetic problems (Grabner and De Smedt, 2012). Against this background, these regions discussed above have been the targets of most transcranial electrical stimulation (tES) studies on arithmetic performance (for a review, see Schroeder et al., 2017), and theta and alpha band ERS/ERD patterns have been used to investigate physiological stimulation effects (Rütsche et al., 2015).

TES comprises different approaches of non-invasive brain stimulation by weak electric currents (generally 1–2 mA), with tDCS being the most commonly used version. In tDCS, a constant direct current is applied via two or more electrodes (anodes and cathodes). It is assumed that the applied current has mainly excitatory effects on the cortical regions beneath the anode, but primarily inhibitory effects on the regions beneath the cathode on the macroscopic level, if stimulation is conducted within a specific conventional range of stimulation intensity and duration (Nitsche et al., 2008; Paulus, 2011). The first studies on the effects of tDCS on arithmetic performance stimulated parietal sites, either unilaterally or bilaterally (Clemens et al., 2013; Hauser et al., 2013; Kasahara et al., 2013; Klein et al., 2013). Clemens et al. (2013) investigated effects of anodal tDCS (a-tDCS) over the right parietal cortex but could not find any effects on multiplication verification tasks. However, an additional fMRI analysis indicated a stronger activation in the AG after stimulation when processing problems which were rehearsed during tDCS. Using a different approach, Klein et al. (2013) found a reduced distractor distance effect in an addition verification task during bilateral parietal a-tDCS, while cathodal stimulation showed no effects. Hauser et al. (2013) applied a-tDCS between two task sessions and found that stimulation over left parietal regions reduced calculation times in large subtractions, while stimulation over right parietal regions as well as bilateral stimulation showed no effects. Interestingly, these effects do not seem to be limited to subtractions, as left anodal / right cathodal parietal tDCS also improved calculation times in complex multiplications (Kasahara et al., 2013). Hence, the left parietal region seems to be a worthwhile target for a-tDCS in the context of arithmetic processing.

The second promising target for tDCS is the left DLPFC. Anodal stimulation of this region has been found to improve arithmetic verification in a group with high math anxiety (Sarkar et al., 2014) and performance in a serial subtraction task (Pope et al., 2015). However, regarding frontal stimulation it is unclear whether the effects of stimulation are domain-specific or if tDCS affects more domain-general functions like working memory (WM) and only indirectly improves arithmetic performance. TDCS, especially over frontal regions, has been found to boost working memory performance (Zaehle et al., 2011; Brunoni and Vanderhasselt, 2014), and Pope et al. (2015) ascribed the positive effects of left frontal a-tDCS on a serial subtraction task to stimulation-induced improvements of working memory. This would also be in line with one study finding beneficial effects of left frontal a-tDCS on a serial addition task only when the stimulation conducted before the task was accompanied by a difficult WM task (Gill et al., 2015). Disentangling the effects of a-tDCS on working memory and arithmetic abilities is complicated, because working memory is an integral part of arithmetic processes, especially procedural calculation (Destefano and LeFevre, 2004; Kasahara et al., 2013). Additionally, like most treatments tDCS does not only have beneficial effects. In a more recent study expanding on the findings of Hauser and colleagues, Rütsche et al. (2015) found that parietal stimulation might indeed enhance performance in large, complex arithmetic problems, but at the same time, this stimulation protocol impaired performance in small, easy problems. This dissociation of effects was accompanied by differential changes in ERS/ERD patterns. While parietal a-tDCS increased lower-alpha ERD during large, complex problems, it decreased theta band ERS during small problems. Hence, there might be a trade-off between beneficial and detrimental stimulation effects, which could prove problematic for the use of a-tDCS as a means to improve arithmetic performance.

This study was performed to expand on prior work by comparing effects of frontal and parietal a-tDCS on arithmetic performance and assessing changes in WM and concomitant EEG. To this end, participants were asked to solve arithmetic problems similar to those used in prior studies (Hauser et al., 2013; Rütsche et al., 2015; small and large additions and subtractions) before, during, and after receiving a-tDCS to either left frontal (targeting the dorsolateral prefrontal cortex; DLPFC) or left parietal regions (targeting the posterior parietal cortex; PPC), or sham stimulation. Additionally, high-density EEG was recorded concomitantly to investigate stimulation-induced changes in ERS/ERD patterns. A short WM task was administered in each phase to investigate stimulation-induced changes in WM performance and to examine task-specificity of stimulation. Based on prior results, we expected that both active stimulations (left frontal and left parietal a-tDCS) improve procedural calculation, with especially frontal stimulation also boosting WM performance (Zaehle et al., 2011; Hauser et al., 2013; Rütsche et al., 2015). This should be accompanied by changes in ERS/ERD patterns in theta and alpha bands, whereby we expected improvements in procedural calculation to be linked to a reduced ERD in alpha bands. Finally, we were eager to investigate, whether the adverse effects of a-tDCS on fact retrieval processes (Rütsche et al., 2015) would replicate in this study.

METHODS

Sample

In total, 72 persons, recruited at the University of Graz and via e-mail and social media, participated in this study. All participants were right handed, without prior or current neurologic or psychiatric disorders, drug use, and any medication potentially influencing the state of their central nervous system. Five participants had to be excluded from analyses because of insufficient performance in at least one block of arithmetic problems (no or only one correct trial). Another two participants had to be excluded because they did not conduct the WM task correctly (no correct trials in at least one block). Finally, three participants had to be excluded from the final analysis because the EEG recorded during the stimulation phase could not be analyzed due to of bad data quality. Hence, the final sample consisted of 62 participants, with 21 receiving left frontal a-tDCS, 20 left parietal a-tDCS and 21 sham stimulation (demographic data listed in **Table 1**). Groups did not differ in arithmetic ability or male/female ratio. However, the group receiving frontal stimulation was significantly younger than the groups receiving parietal or sham stimulation [$F_{(2,59)} = 7.376$; $p = 0.001$]. All participants were thoroughly informed about the study protocol, stimulations and procedures, and gave written informed consent. For participation, they received either 20 € or a study-participation certificate for course credits of 3.5 h. The study was approved by the ethics commission of the University of Graz (No: 60–2015/16).

Arithmetic Tasks

Participants were asked to solve three sets of 64 additions and three sets of 64 subtractions. One of each before, another one of each during, and the final sets after stimulation. Each set consisted of 32 small, easy problems, assumed to be solved by fact retrieval, and 32 large problems, assumed to be solved by procedural calculation. The order of the single items was pseudorandomized. Small additions were one-digit/one-digit problems with a maximum sum of 10. The range of possible operands was 2–8. Small subtractions were constructed by mirroring the small additions ($3 + 6 = 9 \rightarrow 9–6 = 3$). As these rules result in only 24 possible problems, small problems were

TABLE 1 | Demographic data and basic arithmetic ability.

	Total (*N* = 62)		Stimulation groups					
			Frontal (*N* = 21)		Parietal (*N* = 20)		Sham (*N* = 21)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age (years)	25.9	5.1	22.8	3.4	27.5	5.4	27.6	5.0
Arithmetic ability	14.4	3.6	13.7	3.5	14.8	3.4	14.7	4.0
Sex	38 female; 24 male		14 female; 7 male		12 female; 8 male		12 female; 9 male	

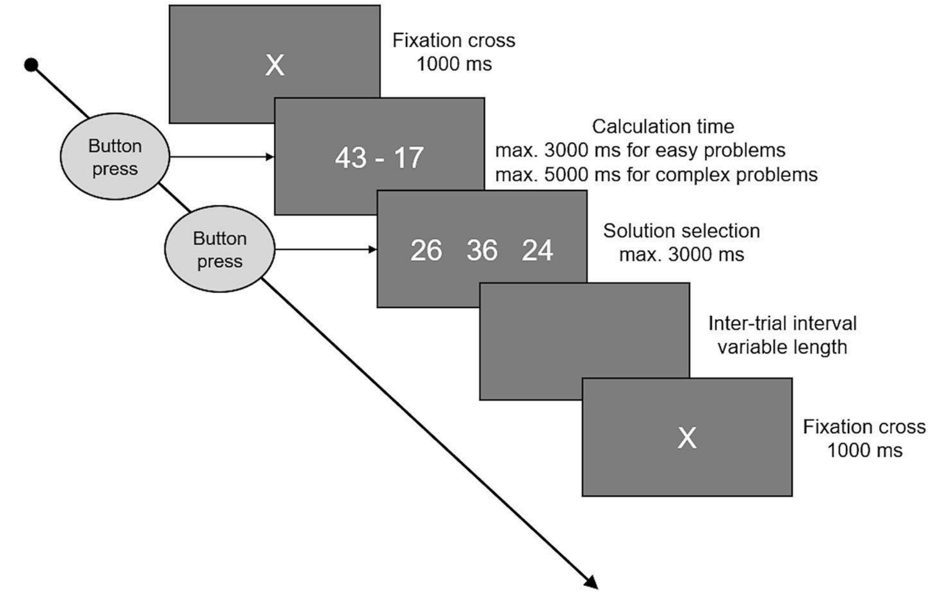


FIGURE 1 | This figure shows the time course of an arithmetic trial. A fixation cross was presented at the beginning, lasting for 1 s. Following the fixation cross, the arithmetic problem was presented on the screen until a button was pressed, indicating that the problem was solved, or time ran out. Maximum time available was 3 s for small and 5 s for large problems. When the button was pressed or time ran out, three options were presented and participants had 3 s to select one of these by pressing the corresponding button. In between the single trials there were inter-trial intervals of variable length, whereby the duration was set to cover the unused time of problem presentation and solution selection, keeping the total trial duration (fixation cross to fixation cross) constant.

the same in all three sets, and in each set eight problems were presented twice. Thereby, the eight repeated problems where different in every set, so that no problem was presented more than four times in total. Large additions were two-digit/two-digit additions with carry, addends between 12 and 59, and sums below 100. Again, the subtractions were constructed by mirroring the additions. Excluding problems with round numbers (e.g., 30) and tie problems, there was a set of 143 large additions/subtractions and 32 of each were randomly selected for each set, without any repetitions. A typical trial is depicted in **Figure 1**. Every trial started with a fixation cross for 1 s, immediately followed by the arithmetic problem. Problems were presented on screen until the participant pressed a button, indicating she/he had solved the problem. The maximum presentation time (time-out) was three (small problems) or five (large problems) seconds, respectively. After the button was pressed or time ran out, participants had 3 s to choose the correct solution from three options by pressing the corresponding button. To keep trial and set durations constant, a blank screen was presented for the time left from the maximal calculation time and maximal solution selection time before the next trial started. After each set, participants could indicate if they used fact retrieval or procedural strategies more often when processing small, easy problems and separately for large, complex problems. To this end, participants were asked to locate a cursor on a bar ranging from "retrieved" (indicating a 100% retrieval rate) to "calculated" (indicating a 100% procedural calculation rate).

The markers for arithmetic performance were accuracy and calculation times. Accuracy was defined as the percentage of trials solved correctly and in time (before the 3 or 5 s time out) in relation to the total number of trials of this type per set. Calculation time was assessed from start of the problem presentation until the button press, indicating that the participant had solved the problem. Mean calculation times per set (block 1–3 and addition vs. subtraction), and difficulty (small vs. large) were calculated, whereby all incorrect trials or problems not solved in time (i.e., no button press before time out) were excluded.

Working Memory

Working memory was assessed by a letter 2-Back task. Task duration was 180 s with a presentation duration of 500 ms for every letter and 1,500 ms of blank screen between to letters (see **Figure 2** for a depiction of the 2-back task). Hence, a single trial was 2,000 ms long, and each run of the 2-back task consisted of 90 trials. The letters used were "A," "B," "C," "D," and "E," which appeared in pseudorandomized order to achieve 60 non-target and 30 target trials. Target trials were defined as trials in which the presented letter matched the letter from two trials ago. Participants were instructed to indicate target trials by pressing a button and could do so during the whole trial duration from presentation of one letter to the presentation of the next (500 ms presentation time plus 1,500 ms blank screen). They had to refrain from pressing any button during non-target trials. A correct reaction (CR) consisted of a button press during a target trial, a correct rejection (CRJ) of refraining from a button press during non-target trials. A false alarm (FA) occurred when the button was pressed during a non-target trial and a miss (M) was defined as the absence of a button press during a target trial. Working memory performance was assessed by reaction time (WM-RT) over all correct trials and by accuracy (WM-ACC), calculated by WM-ACC = [1–((FA + M) / 90)) * 100].

Basic Arithmetic Ability

The subtest "Rechenzeichen" (arithmetic operators) of the IST-2000R (Liepmann et al., 2007) was used as a short assessment of participants' basic arithmetic ability. In this subtest, participants are presented with 20 items consisting of an arithmetic problem and its solution but without the operators (e.g., A ? B ? C = D). They have 10 min to identify the correct operators for the 20 problems. The number of correctly solved problems represents the raw score in this subtest.

Transcranial Direct Current Stimulation

Stimulation was applied utilizing a NeuroConn DC-Stimulator Plus (NeuroConn, Ilmenau, Germany). Electrodes were rectangular rubber electrodes sized 3 by 3 cm for the anode and 5 by 7 cm for the cathode. The anode was placed over EEG position F3 for frontal stimulation, targeting the left DLPFC, and over P3 for parietal stimulation, targeting the left PPC. The cathode was placed over the contralateral supraorbital site. For half of the participants receiving sham stimulation the electrodes were mounted as for frontal stimulation and for the other half as for parietal stimulation. Electrodes were applied directly to the scalp with an about 1–2 mm thick layer of Ten20 paste (Weaver and Company, Aurora, USA) and held in place by the EEG cap mounted above the stimulation electrodes. In the active stimulation groups, tDCS was applied for 25 min with an intensity of 1 mA and fade in/out phases of 30 s in which the current was slowly ramped up/down. Current density under the electrode was 0.11 mA/cm^2 for the anode and 0.03 mA/cm^2 under the cathode. Sham stimulation consisted of a 30 s fade in phase followed by 50 s of applied current (1 mA) and 30 s of fade out, in order to induce the same sensory perception as the active stimulations. Impedances were comparable between the three groups [frontal: M = 3.41 kΩ (SD = 2.32); parietal: M = 3.68 kΩ (SD = 2.11); sham M = 4.39 kΩ (SD = 2.19); $F_{(2,59)} = 1.208$; $p = 0.306$; $\eta_p^2 = 0.039$]. Furthermore, the stimulation was applied in a double-blind way by using the study mode of the DC-Stimulator Plus. Here, a code list was prepared with one code for each subject, and entering the respective code either started active or sham stimulation.

Electroencephalography

Recording

The EEG recording was conducted while the participants processed the arithmetic and working memory tasks in a separate, normally lit and quiet room, using a 64-channel BioSemi ActiveTwo EEG system (BioSemi, Amsterdam, Netherlands). Electrodes were mounted according to the 10:20 system (Jasper, 1958) using BioSemi head caps and Signagel (Parker Laboratories, Fairfield, USA) to ensure appropriate contact. As the tDCS electrodes were mounted at the positions F3, P3, Fp2, and AF8 these EEG electrodes were not mounted.

Preprocessing

Data was analyzed using MNE (Gramfort et al., 2013, 2014) and additional, custom-built Python code. For the EEG recordings before and after stimulation, pre-processing was done semi-automatically using an average reference, a 1 Hz high-pass filter and visual inspection regarding prominent artifacts and bad channels before applying an independent component analysis (ICA) to remove ocular artifacts. This was followed by applying a notch (48–52 Hz) and a low-pass filter (60 Hz) before a second visual inspection to detect any remaining artifacts. For the EEG data recorded during the stimulation phase, an additional principal component analysis (PCA) step to remove artifacts induced by the stimulator during sham stimulation (repeated impedance checks) was applied before the ICA was performed. Finally, data was prepared separately for each frequency band of interest (theta 3–6 Hz; low alpha 8–10 Hz, and high alpha 10–13 Hz) by applying adequate band-pass filters. The chosen frequency ranges were based on prior studies (Grabner and De Smedt, 2011, 2012). Afterwards, the mean power during the reference interval (R; fixation cross; 1,000 ms) and during activation (A; calculation time from problem onset until button press) in all correct trials consisting of more than 50% artifact-free data was assessed for each frequency band. In each block (before, during, and after stimulation) and arithmetic task (small/large additions/subtractions) on average, $M = 0.53$; $SD = 1.12$ had to be excluded because of artifacts and $M = 28.56$; $SD = 3.96$ trials were used for analysis. The mean power during reference and activation intervals was averaged over all used trials (separately for additions and subtractions as well as for small and large problems), and ERS/ERD values for the four types of arithmetic problems and the three phases were calculated by ERS/ERD = ((A–R) / R) * 100. Hence, positive values indicate ERS (an increase of power from the reference interval to activation) and negative values indicate ERD (a decrease of power from reference interval to activation). Finally, the single channels were grouped into clusters, and single-channel ERS/ERD values were averaged to result in a single ERS/ERD value for each cluster, problem type, and time point (before, during, and after stimulation). The clusters used were left frontal (Fp1, AF3, AF7, F7, FC5, FC3, and FC1), right frontal (AF4, F6, F4, F2, FC6, FC4, FC2), left parietal (CP5, CP3, CP1, P7, PO7), and right parietal (CP2, CP4, CP6, P8, P6, P4, P2, PO8, PO4). In addition to the channels not mounted because of the tDCS electrodes, the channels F1, F5, F8, P1, P5, and PO3 were excluded for all participants, as these channels were closest to the stimulation electrodes and did not yield processable EEG in most of the participants receiving active stimulation. **Figure 3** depicts the procedure of EEG data processing.

Procedure

The experimental session consisted of three parts (see **Figure 4**). In the beginning, participants were asked to answer a demographic questionnaire, followed by a short test to ascertain right hand dominance (HDT; Steingrüber and Lienert, 1971), a verbal fluency test (RWT; Aschenbrenner et al., 2000) and the Comprehensive-Trial-Making- Test (CTMT; Reynolds, 2002). The latter two tests were not used in this study, but were

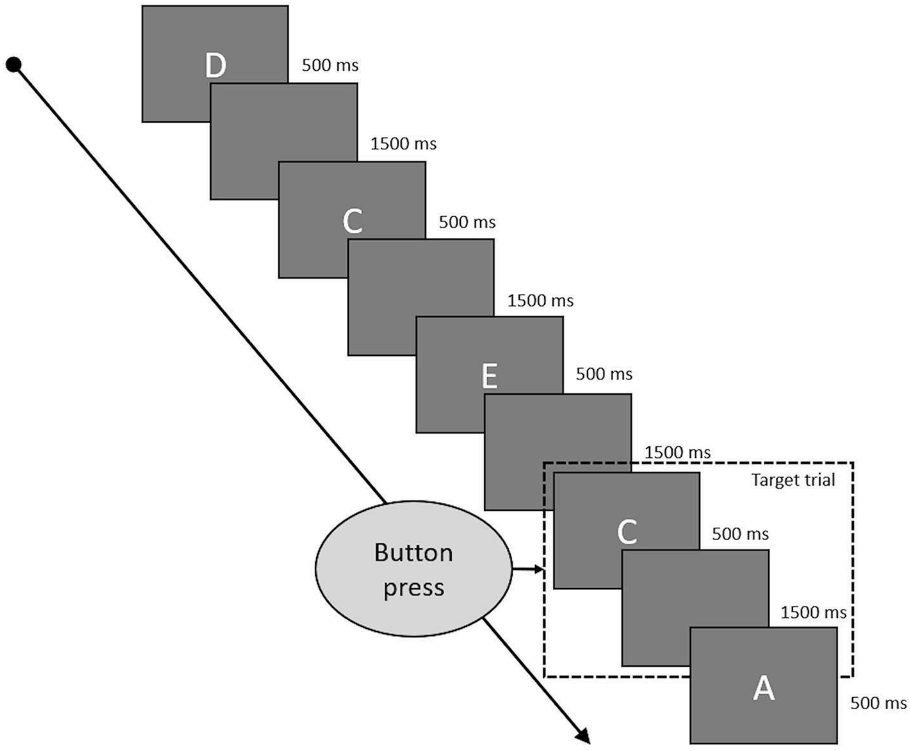


FIGURE 2 | This figure shows some typical trials of the 2-back task. Every letter was presented for 500 ms and followed by a blank screen for 1,500 ms. Target trials were the letters which were identical to the letter presented two trials before. Participants had to press a button if the current trial was a target trial and could do so while the letter was presented to them and during the following blank screen phase.

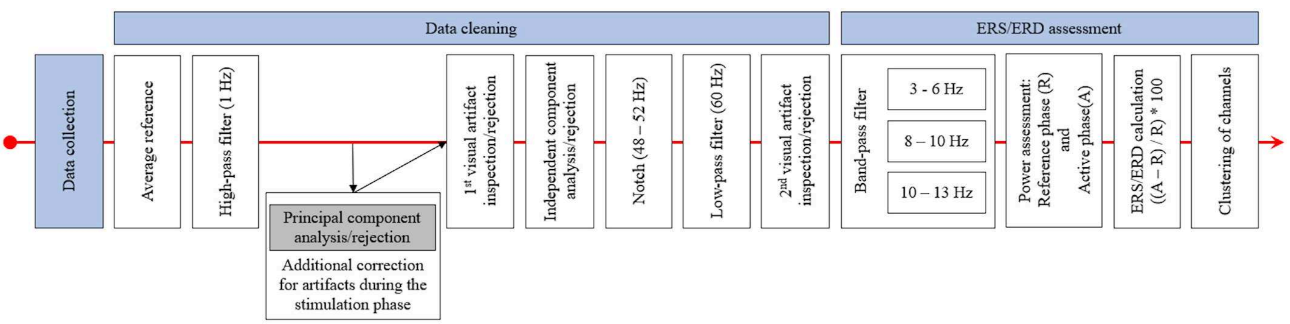


FIGURE 3 | This figure displays the procedure of EEG data pre-processing and ERS/ERD calculation. After data collection, the EEG data was re-referenced and high-pass filtered. This was followed by a principal component analysis for the data recorded during stimulation and by the first visual inspection and rejection of artifacts. The independent component analysis was applied after this first inspection and after bad components were excluded the EEG data was notch and low-pass filtered. The final step was the second visual inspection and artifact rejection. ERS/ERD calculation started with band-pass filtering of the EEG to only include data in the frequency band of interest. Based on this data, the band power in the reference and active intervals was assessed and ERS/ERD values were calculated. The final step was the clustering, by averaging ERS/ERD values over the channels included in the different clusters.

part of another project. The last test conducted before the EEG and tES electrodes were mounted was the *"Rechenzeichen" (operators)*-part of the Intelligenz-Struktur-Test 2000R (IST-2000R; Liepmann et al., 2007) as a short assessment of basic arithmetic abilities. Following this, the EEG and tDCS electrodes were mounted and the main test session with EEG recording and stimulation took place. This main test session was the second part of the study and consisted of three blocks, with two sets of arithmetic problems (one containing additions and one subtractions) and one WM task (N-Back task) each. The running order of the arithmetic sets and the WM task was pseudorandomized in order to be balanced over subjects and groups, but was constant for all three blocks (before, during, and after stimulation) within each person (e.g., if for a participant the

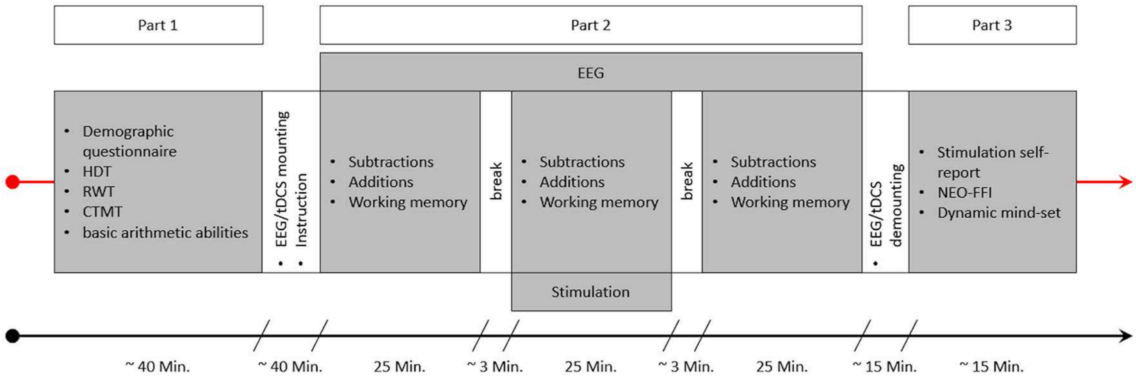


FIGURE 4 | This figure depicts an experimental session. After getting informed about the experiment and signing the consent form the first part of the study started, consisting of a demographic questionnaire, the HDT, RWT, and CTMT, as well as the "Rechenzeichen" (operators) scale of the IST serving as a measure of basic arithmetic ability. This part was followed by the specific instructions for the arithmetic and working memory tasks and the mounting of the EEG and tDCS electrodes. Part 2 consisted of the three blocks of arithmetic and working memory tasks. The order of the three tasks varied between the subjects in a pseudorandomized manner. Stimulation was applied for 25 min during block 2, and EEG was recorded during all three blocks. After part 2 was finished, the EEG and tDCS electrodes were demounted and participants could wash their hair. Finally, in part 3, participants were asked to complete a stimulation self-report, the NEO-FFI, and the dynamic-mindset questionnaire.

order was working memory, additions, subtractions in block 1 the tasks would be in the same order in blocks 2 and 3 for this person). The arithmetic sets and the WM task were separated by breaks of 50 s, with longer breaks between the three blocks in order to start and stop stimulation. Before the start of an arithmetic set or the WM task, participants were informed by a sound signal and a message on screen that the tasks will resume, and another message informing them about which task will be next. After the main part was finished, EEG and tDCS electrodes were demounted and participants were asked whether they think they received active or sham stimulation. In the final part, they were asked to answer a stimulation self-report as well as the German version of the NEO-Five-Factor-Inventory (NEO-FFI; Borkenau and Ostendorf, 2008), and a questionnaire regarding their dynamic mind-set, both being part of another project.

Statistical Analysis

All analyses were conducted using SPSS 25 (IBM, Armonk, USA). Stimulation-induced changes in arithmetic and working memory performance were analyzed using mixed design ANCOVAs with within-subjects factor time (before, during, and after stimulation) and the between-subjects factor treatment (frontal a-tDCS, parietal a-tDCS, and sham stimulation). Analyses were calculated separately for accuracy and calculation time of each arithmetic type (small additions, large additions, small subtractions, large subtractions), and for overall accuracy and reaction time of the WM task. Changes in ERS/ERD values were analyzed by calculating mixed design ANCOVAs with the within-subjects factors time (before, during, and after stimulation), and location (frontal, parietal), and the between-subjects factor treatment (frontal a-tDCS, parietal a-tDCS, and sham stimulation). These analyses were carried out separately for each frequency band and hemisphere, but only for these types of arithmetic problems in which stimulation changed performance, as the main research question regarding the EEG was whether stimulation-induced behavioral changes are reflected in ERS/ERD patterns. Participants' age, sex, and basic arithmetic abilities were used as covariates in all analyses concerning arithmetic performance and related EEG. For analysis of working memory performance, only age and sex were used as covariates. Greenhouse-Geisser correction was used if sphericity could not be assumed as indicated by a significant Mauchly's test of sphericity. Efficacy of blinding was analyzed using a chi-square test on the stimulation self-report data.

RESULTS

Behavioral

Small Additions

For accuracy in small additions (overall $M = 98.00\%$; $SD = 1.70$), the ANCOVA showed no significant main effects of time [$F_{(2,112)} = 1.927$; $p = 0.150$; $\eta_p^2 = 0.033$] or treatment [$F_{(2,56)} = 0.157$; $p = 0.855$; $\eta_p^2 = 0.006$] nor a significant interaction time * treatment [$F_{(4,112)} = 0.146$; $p = 0.965$; $\eta_p^2 = 0.005$].

For calculation times ($M = 0.75$ s; $SD = 0.13$), the ANCOVA showed a significant main effect of time [$F_{(2,112)} = 5.792$, $p = 0.004$; $\eta_p^2 = 0.094$] but no significant effect of treatment [$F_{(2,56)} = 2.583$; $p = 0.085$; $\eta_p^2 = 0.084$] or interaction time * treatment [$F_{(4,112)} = 0.471$; $p = 0.757$; $\eta_p^2 = 0.017$]. Pairwise comparisons showed, that calculation times before treatment ($M = 0.77$ s, $SD = 0.17$) were slower than during treatment ($M = 0.74$; $SD = 0.14$; $p = 0.009$) and after treatment ($M = 0.73$; $SD = 0.14$; $p = 0.006$). Calculation times during and after treatment did not differ ($p = 0.552$).

Large Additions

For accuracy in large additions ($M = 80.63\%$; $SD = 8.62$), the ANCOVA showed neither significant main effects of time [$F_{(2,112)}$

$= 0.215$; $p = 0.807$; $\eta_p^2 = 0.004$] and treatment [$F_{(2,56)} = 0.226$; $p = 0.798$; $\eta_p^2 = 0.008$] nor a significant interaction time * treatment [$F_{(4,112)} = 1.633$; $p = 0.171$; $\eta_p^2 = 0.055$].

Similarly, for calculation times ($M = 2.52$ s; $SD = 0.66$), the ANCOVA showed neither significant main effects of time [$F_{(1.74,97.60)} = 0.419$; $p = 0.631$; $\eta_p^2 = 0.007$] and treatment [$F_{(2,56)} = 0.662$; $p = 0.520$; $\eta_p^2 = 0.023$] nor a significant interaction time * treatment [$F_{(3.49,97.60)} = 1.129$; $p = 0.345$; $\eta_p^2 = 0.039$].

Small Subtractions

For accuracy in small subtractions ($M = 96.76\%$; $SD = 2.89$), the ANCOVA showed neither significant main effects of time [$F_{(2,112)} = 0.217$; $p = 0.805$; $\eta_p^2 = 0.004$] and treatment [$F_{(256)} = 1.089$; $p = 0.344$; $\eta_p^2 = 0.037$] nor a significant interaction time * treatment [$F_{(4,112)} = 0.464$; $p = 0.762$; $\eta_p^2 = 0.016$].

Similarly as for small additions, the ANCOVA for calculation times in small subtractions ($M = 0.84$ s; $SD = 0.20$) also showed a significant main effect of time [$F_{(2,112)} = 5.837$, $p = 0.004$; $\eta_p^2 = 0.094$] but no significant effect of treatment [$F_{(2,56)} = 1.958$; $p = 0.151$; $\eta_p^2 = 0.065$] or time * treatment interaction [$F_{(4,112)} = 0.694$; $p = 0.597$; $\eta_p^2 = 0.024$]. Pairwise comparisons showed, that calculation times before stimulation ($M = 0.87$ s, $SD = 0.21$) were slower than during stimulation ($M = 0.84$; $SD = 0.22$; $p = 0.036$) and after stimulation ($M = 0.82$; $SD = 0.22$; $p = 0.003$). Calculation times during and after stimulation did not differ ($p =0.167$).

Large Subtractions

The ANCOVA showed a significant main effect of time for accuracy ($M = 74.83\%$; $SD = 12.29$) in large subtractions [$F_{(2,112)} = 12.749$; $p < 0.001$; $\eta_p^2 = 0.185$] but no main effect of treatment [$F_{(2,56)} = 0.144$; $p = 0.866$; $\eta_p^2 = 0.005$] or interaction between time and treatment [$F_{(4,112)} = 1.406$; $p = 0.237$; $\eta_p^2 = 0.048$]. Pairwise comparisons showed that accuracy before treatment ($M = 71.47$; $SD = 15.30$) was lower than during treatment ($M = 76.31$; $SD = 10.77$; $p < 0.001$) and after treatment ($M = 76.71$; $SD = 13.74$; $p < 0.001$), while there was no difference between during and after treatment ($p = 0.708$).

For calculation times in large subtractions ($M = 2.83$ s; $SD = 0.65$), the ANCOVA showed a significant main effect of time [$F_{(2,112)} = 5.786$; $p = 0.004$; $\eta_p^2 = 0.094$) but not for treatment [$F_{(2,56)} = 0.096$; $p = 0.909$; $\eta_p^2 = 0.003$]. However, there was a significant interaction of time * treatment [$F_{(4,112)} = 2.787$; $p = 0.030$; $\eta_p^2 = 0.091$]. Pairwise comparisons showed that the interaction was driven by differences in calculation time reductions over time between the treatment groups. The group receiving frontal a-tDCS showed a significant reduction of calculation times from before treatment ($M = 2.99$; $SD = 0.75$) to during treatment ($M = 2.81$; $SD = 0.72$; $p = 0.013$) and after treatment ($M = 2.76$; $SD = 0.79$; $p = 0.013$) while there was no difference in calculation times between the blocks during and after treatment ($p = 0.891$). The group receiving sham stimulation, on the other hand, showed no reduction of calculation time from before treatment ($M = 2.93$; $SD = 0.50$) to during treatment ($M = 2.91$; $SD = 0.55$; $p = 0.997$), but in the block after treatment ($M = 2.74$; $SD = 0.56$) they were significantly faster than before ($p = 0.008$) and during treatment ($p = 0.002$). Finally, the group receiving parietal a-tDCS showed no differences in calculation times over time (before; $M = 2.78$; $SD = 0.86$; during; $M = 2.75$; $SD = 0.72$; after: $M = 2.81$; $SD = 0.62$; all $p > 0.05$). Results are depicted in **Figure 5**, and calculation time changes on a single person level are given as additional information in **Figure SM1**.

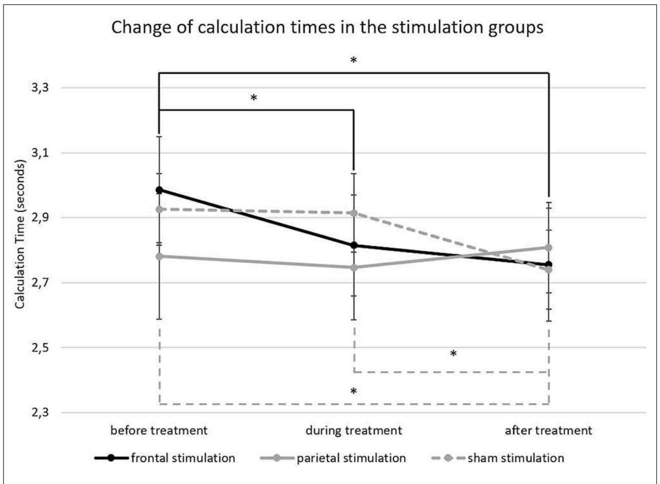


FIGURE 5 | This figure depicts the change of calculation time in large subtractions from before to after treatment. The error bars represent the standard error of the mean and * indicates significant differences (unbroken lines depict differences in the frontal stimulation group and dashed lines in the sham group).

Working Memory

Regarding WM accuracy, the ANCOVA showed a significant main effect of time [$F_{(2,114)} = 5.929$; $p = 0.004$; $\eta_p^2 = 0.094$] but no effect of treatment ($F(2,57) = 0.320$; $p = 0.727$; $\eta_p^2 = 0.011$) or time * treatment interaction [$F_{(4,114)} = 1.678$; $p = 0.160$; $\eta_p^2 = 0.056$]. Pairwise comparisons showed that the accuracy before treatment ($M = 89.23$; $SD = 5.85$) was lower than the accuracy during ($M = 90.79$; $SD = 6.99$; $p = 0.035$) and after treatment ($M = 91.69$; $SD = 7.02$; $p = 0.004$) but did not differ between during and after treatment ($p = 0.155$).

Similarly, for WM reaction times, the ANCOVA showed a significant main effect of time [$F_{(1.767,100.698)} = 3.353$; $p = 0.045$; $\eta_p^2 = 0.056$], but no effect of treatment [$F_{(2,57)} = 0.828$; $p = 0.442$; $\eta_p^2 = 0.028$] or time * treatment interaction [$F_{(3.533;100.698)} = 0.299$; $p = 0.857$; $\eta_p^2 = 0.010$]. Pairwise comparisons showed that the reaction times after treatment ($M = 0.57$; $SD = 0.13$) were faster than during treatment ($M = 0.60$; $SD = 0.13$; $p = 0.002$). However, reaction times before treatment ($M = 0.59$; $SD = 0.13$) lay in between the times achieved during and after treatment and did not significantly differ from either (both $p > 0.05$).

ERS/ERD

Theta Band ERS/ERD During Large Subtractions

Main results are depicted in **Figure 6**; additional topographic information containing single channel information is displayed

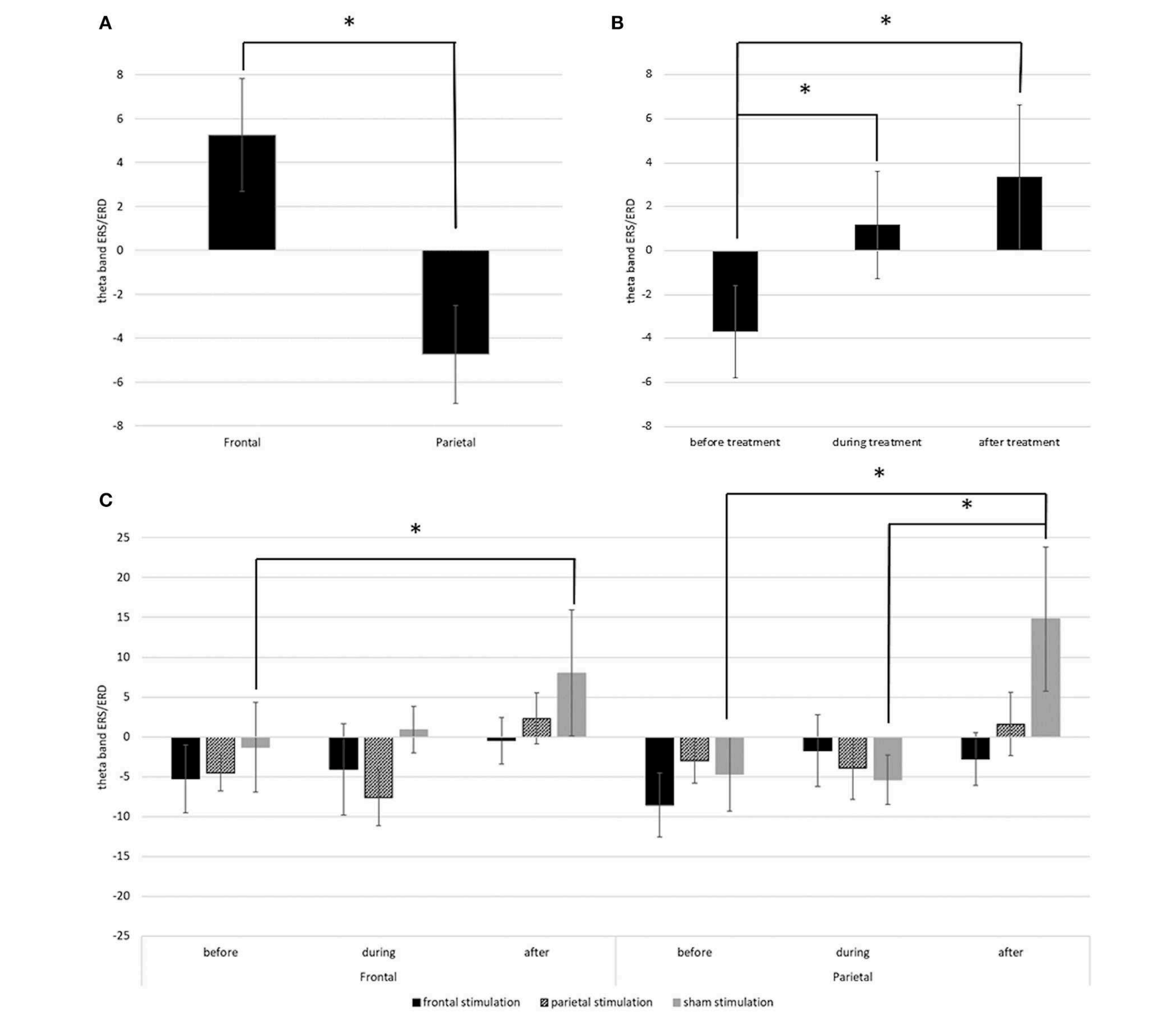


FIGURE 6 | (A) Depicts the mean theta band ERS/ERD values in the left frontal and parietal regions over all time points. **(B)** Depicts the overall change of theta band ERS/ERD in the left hemisphere from before to after stimulation. **(C)** Depicts the change of theta band ERS/ERD values in right frontal and right parietal regions in the different treatment groups from before to after stimulation. Error bars represent the standard error of the mean and * indicates significant differences.

in **Figure 7**. For the left hemisphere, the ANCOVA showed a significant main effect of time [$F_{(1.437,80.492)} = 4.185$; $p = 0.030$; $\eta^2_p = 0.070$] and location [$F_{(1,56)} = 36.514$; $p < 0.001$; $\eta^2_p = 0.395$] but no effect of treatment [$F_{(2,56)} = 0.655$; $p = 0.524$; $\eta^2_p = 0.023$] or time * location [$F_{(1.603,89.782)} = 2.924$; $p = 0.070$; $\eta^2_p = 0.050$], time * treatment [$F_{(2.875,80.492)} = 0.506$; $p = 0.671$; $\eta^2_p = 0.018$], location * treatment [$F_{(2,56)} = 1.127$; $p = 0.331$; $\eta^2_p = 0.039$], and time * location * treatment interactions [$F_{(1.603,89.782)} = 1.771$; $p = 0.140$; $\eta^2_p = 0.059$]. Overall, participants showed a theta band ERS over frontal regions ($M = 5.26$; $SD = 20.26$), but an ERD over parietal regions ($M = -4.74$; $SD = 17.50$) with a decrease of ERD / increase of ERS over time (**Figures 5A,B**). Thereby, before treatment, participants showed a theta band ERD ($M = -3.69$; $SD = 16.56$) which was significantly different from the values during ($M = 1.16$; $SD = 19.23$; $p = 0.042$) and after treatment ($M = 3.31$; $SD = 26.14$; $p < 0.001$) where participants showed ERS patterns. ERS values during and after treatment did not differ ($p = 0.494$).

For the right hemisphere, the ANCOVA showed significant main effect of time [$F_{(1.574,88.162)} = 7.070$; $p = 0.003$; $\eta^2_p = 0.112$] and a significant time * location * treatment interaction [$F_{(3.918,109.692)} = 3.293$; $p = 0.014$; $\eta^2_p = 0.105$]. Effects of location [$F_{(1,56)} = 0.017$; $p = 0.898$; $\eta^2_p = 0.000$] and treatment [$F_{(2,56)} = 0.752$; $p = 0.476$; $\eta^2_p = 0.026$] and the interaction time * location [$F_{(1.959,109.692)} = 0.668$; $p = 0.515$; $\eta^2_p = 0.012$], time * treatment [$F_{(3.148,88.162)} = 0.901$; $p = 0.448$; $\eta^2_p = 0.031$], and

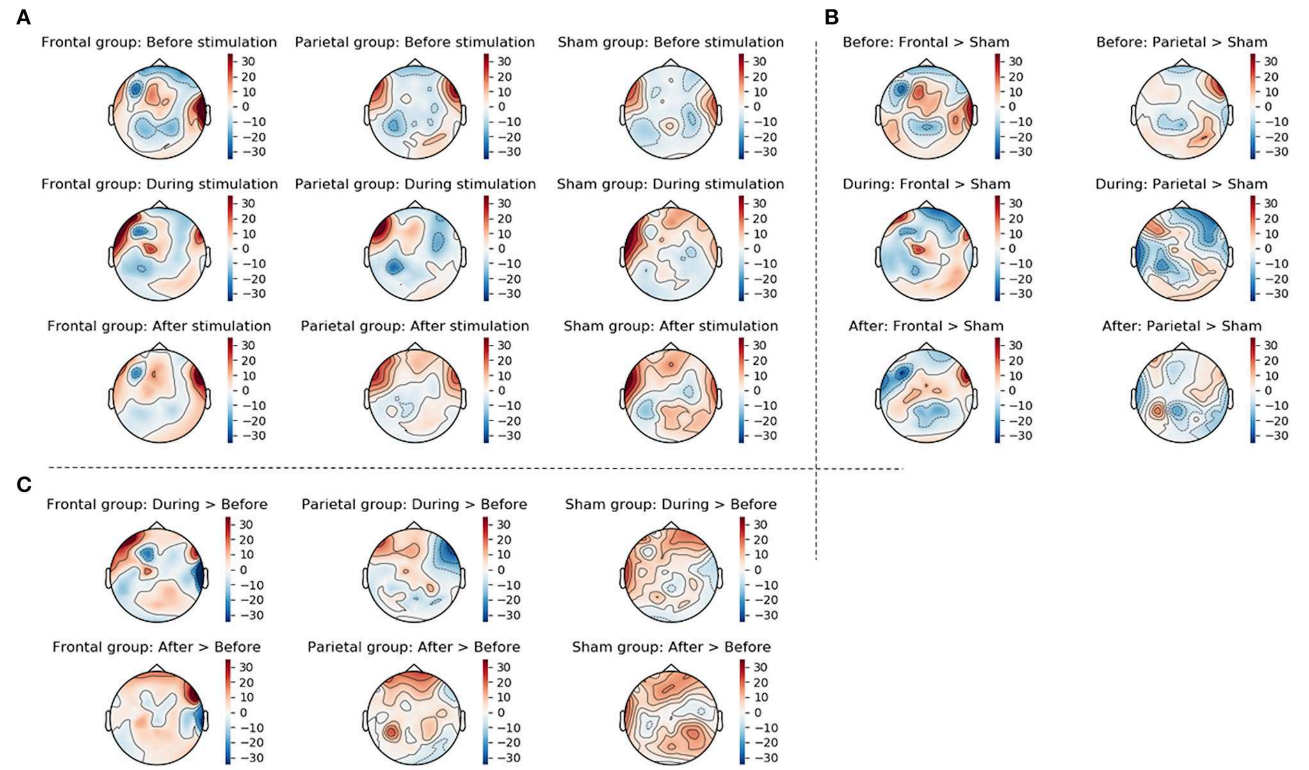


FIGURE 7 | This figure displays the topographic information of average ERS/ERD values in the three groups (frontal stimulation parietal stimulation, sham stimulation) before, during, and after stimulation **(A)**, the difference between frontal and sham groups as well as the difference between parietal and sham groups at the three time points **(B)**, and the change from before to during stimulation and from before to after stimulation in the three groups **(C)**. Changes were calculated by subtracting ERS/ERD values in the sham group from those in the actively stimulated groups for the between group changes displayed in **(B)** and by subtracting the ERS/ERD values before stimulation from those during or after stimulation for the within group changes displayed in **(C)**.

location * treatment [$F_{(2,56)} = 0.272$; $p = 0.763$; $\eta_p^2 = 0.010$] were not significant. Pairwise comparisons showed, that the interaction time * location * treatment was driven by the sham stimulated group. In this group, there was a significant difference in ERS/ERD values over frontal regions between before ($M = -0.84$; $SD = 36.54$), and after treatment ($M = 7.53$; $SD = 42.68$; $p = 0.033$) while the ERS/ERD values in the active stimulation conditions did not differ between the timepoints (both $p > 0.05$). Over parietal regions, the sham group also showed a significant change in ERS/ERD values from before ($M = -4.46$; $SD = 31.59$) to after treatment ($M = 14.80$; $SD = 49.39$; $p < 0.001$), and additional differences between the ERS/ERD values during treatment ($M = -5.35$; $SD = 31.79$) and after treatment ($p = 0.001$) but no difference between ERS/ERD values before and during treatment ($p = 0.829$). All other pairwise comparisons were non-significant ($p > 0.05$).

Low Alpha Band ERS/ERD During Large Subtractions

For the left hemisphere, the ANCOVA showed a significant effect of locations [$F_{(1,56)} = 36.806$; $p < 0.001$; $\eta_p^2 = 0.397$] with a stronger ERD in parietal regions ($M = -31.63$; $SD = 21.26$) than in frontal regions ($M = -13.68$; $SD = 35.23$). No other main effects or interactions were significant (all $p > 0.05$).

Similarly, for the right hemisphere, the ANCOVA also only showed a significant effect of location [$F_{(1,56)} = 10.729$; $p = 0.002$; $\eta_p^2 = 0.161$], with stronger ERD in parietal regions ($M = -26.49$; $SD = 29.27$) than in frontal regions ($M = -21.26$; $SD = 27.66$). All other main effects and interactions were non-significant (all $p > 0.05$).

High Alpha Band ERS/ERD During Large Subtractions

For high alpha, the ANCOVA also showed a significant main effect of location in the left hemisphere [$F_{(1,56)} = 16.414$; $p < 0.001$; $\eta_p^2 = 0.227$], with a stronger ERD in parietal regions ($M = -26.16$; $SD = 17.77$) as compared to frontal regions ($M = -17.89$; $SD = 22.28$). Again, all other main effects and interactions were non-significant (all $p > 0.05$).

This also holds true for the right hemisphere, were, again, only the main effect of location proved significant [$F_{(1,56)} = 18.199$; $p < 0.001$; $\eta_p^2 = 0.245$], with a stronger ERD in parietal regions ($M = -24.48$; $SD = 19.71$) than in frontal regions ($M = -18.37$; $SD = 21.13$). All other main effects and interactions were non-significant (all $p > 0.05$).

Blinding

A chi-square test showed no relation between the subjective perception and actually applied stimulation [active vs. sham; $\chi^2(2, N = 62) = 2.357$; $p = 0.308$]. In the frontally stimulated group, there were 13 participants assuming to have received active stimulation and 8 assuming to have received sham stimulation. In the parietally stimulated group this ratio was 8–12 and in the sham group 9–12.

DISCUSSION

The aim of the present study was to extend prior research by directly comparing the effects of a-tDCS over left frontal (targeting the DLPFC) and left parietal (targeting the PPC) regions on arithmetic performance in small and large problems of different operations and on EEG activity. In order to be able to conduct a more fine-grained analysis, performance and EEG were assessed before, during, and after stimulation, allowing for a separation of online and after-effects of stimulation on both, the behavioral and the neurophysiological level. Additionally, WM performance was assessed to investigate whether the effects of a-tDCS on arithmetic performance are task specific. Overall, we found no general tDCS related improvements in arithmetic or working memory performance, but there is some evidence for an acceleration of training gains in participants receiving frontal stimulation. These participants showed a significant improvement in calculation times in large subtractions from before to during and after stimulation, while participants receiving sham stimulation showed a similar change only in the last block. However, this admittedly small effect was not reflected in ERS/ERD patterns.

Behavioral Effects

Contrary to the expectations and results in prior literature (Hauser et al., 2013; Pope et al., 2015; Rütsche et al., 2015), neither left frontal, nor left parietal a-tDCS induced a general improvement in arithmetic performance. While participants did show some performance improvements over time in all but one type of arithmetic problems (in large additions no improvements emerged), these were mostly general improvements in accuracy (large subtractions) or calculation times (small additions) that can be attributed to practice effects. However, in large subtractions, left frontal a-tDCS led to an accelerated improvement in calculation times from before to during stimulation, as compared to the group receiving sham stimulation. The group receiving sham stimulation also improved, but later on, and the group receiving a-tDCS over parietal regions showed no improvement in calculation times over time. The earlier reduction of calculation times in the frontally stimulated group might indicate accelerated training gains. Although large problems were not repeated, the procedural strategies used to solve them have been trained over the course of the three arithmetic blocks, as can be seen from the general improvement in accuracy and the gains in calculation times in the frontally and sham stimulated groups. Stimulation effects might have not been strong enough for an overall performance improvement, but sufficient to support training gains. This is in line with a recent meta-analysis showing that tDCS effects on training or learning gains are generally larger than those on performance (Simonsmeier et al., 2018).

The absence of any improvement in the group receiving left parietal a-tDCS is startling, as it was parietal a-tDCS that showed promising effects on calculation times in large arithmetic problems (Rütsche et al., 2015) and specifically large subtractions (Hauser et al., 2013) in earlier studies. On the other hand, the detrimental effects of parietal a-tDCS on the accuracy in small arithmetic problems reported by Rütsche et al. (2015) also could not be replicated. All groups showed very fast calculation times and high accuracies in small additions as well as small subtractions and performance improved over time without any stimulation induced differences. One possible explanation for these diverging results could be found in the difference of the timing of tDCS between this study and prior studies finding positive effects (Hauser et al., 2013; Pope et al., 2015; Rütsche et al., 2015). In these prior studies, stimulation was applied before or in between two sets of the arithmetic tasks, while in the present study stimulation was applied concomitant to the task. There is some evidence for differences in stimulation effects depending on preexisting activity, indicating a neuronal state dependency of non-invasive brain stimulation effects in general (Silvanto et al., 2008; Romei et al., 2016). Additionally, there are other slight differences to prior studies like electrode size (smaller in the present study), positioning of the anode (P3 in the present study but P5/CP5 in Rütsche et al., 2015), or stimulation intensity (1 mA in the present study but 2 mA in Pope et al., 2015), which might have contributed to the differences in results. Furthermore, calculation times were faster in the present study as compared to prior ones (Hauser et al., 2013; Rütsche et al., 2015). This might have been brought about by the time limits for calculations (3 sec. for small; 5 sec. for large problems) which were, together with the inter-trial intervals, implemented to keep the set durations constant. These limits could have induced some time pressure leading to a faster processing of the tasks and hence, less scope for further improvements by stimulation, especially in large problems. Finally, the question regarding the task specificity of tDCS, whether arithmetic processes are improved directly or indirectly by beneficial effects of tDCS on working memory, remains ambiguous. Not only were tDCS effects on arithmetic performance limited to one task and rather small, but, contrary to prior results (Brunoni and Vanderhasselt, 2014), results show no tDCS related effects on working memory.

Electrophysiological Effects

Since on a behavioral level there only was a small stimulation effect on training gains in large subtractions, the focus regarding neurophysiological changes was set to ERS/ERD patterns emerging during this type of problem. However, results showed no clear stimulation related effects. Interestingly, there was no change in low and high alpha ERD patterns in general, although there was a general improvement in performance in large subtractions.

There were, however, some interesting changes in theta band ERS/ERD patterns accompanying the processing of large subtractions over time. In the left hemisphere, there was a general change from an ERD pattern during block 1 to an ERS pattern during block 3. As in the area of mental arithmetic theta band

ERS has been associated with the cognitive less demanding fact retrieval process (De Smedt et al., 2009; Grabner and De Smedt, 2011) and fact training led to an increase in theta band ERS (Soltanlou et al., 2019), this general increase in theta band ERS could reflect the training effect. However, as in this study, the procedural problems were never repeated and hence, no fact training existed, it could only reflect a decrease in cognitive demand of the increasingly trained procedural calculation process. Another probably more plausible explanation could be that this change reflects an increasing demand on attentional processes and cognitive control. Several studies found that frontal theta ERS is also associated with these processes (Missonnier et al., 2006; Cavanagh and Frank, 2014; Ishii et al., 2014). All three blocks consisted of a WM part and two sets of arithmetic problems, which were only separated by short breaks. As the duration increases, the demand on attentional and control processes to carry out the tasks could have increased and hence, led to an increase in associated theta band ERS. The stronger ERS in frontal regions as compared to parietal regions supports this notion.

A similar pattern was also found in the right hemisphere, but only in the group receiving sham stimulation. While this could be a fortuitous effect, it also could indicate an effect of stimulation. Anodal tDCS has been shown to induce wide spread effects and modulate activity in broad networks and different sites of the brain (Polanía et al., 2011; Pena-Gomez et al., 2012). In this case, anodal stimulation of left hemispheric sites of the brain could have modulated activity in right hemispheric regions via mechanisms of interhemispheric inhibition and hence, might have hindered a similar theta band ERS increase as seen in the sham group. However, this explanation can so far only be speculative, and further research is needed to investigate such effects, especially as there were no effects in the stimulated sites themselves.

Another possible explanation could be that the absence of theta band ERS changes in the right hemisphere of the stimulated groups is caused by the cathodal return electrode. This is also one of the limitations of this study. The cathode was mounted at the contralateral supraorbital site. Although a larger electrode (5×7 cm) was used, rendering the applied current density beneath it rather low (0.03 mA/cm^2), an inhibitory effect of this electrode on right frontal areas, especially the frontopolar area, cannot be ruled out completely. This could have disturbed the theta band ERS change in the right hemisphere, at least in frontal regions. However, both explanations come short in explaining why both stimulated groups (left frontally and left parietally) show an absence of theta band ERS increase in right frontal and right parietal sites as compared to the sham group. Another possible issue brought about by the cathode is its potential impact on behavioral effects. As frontopolar regions have been thought of as a metacognitive hub-region (Burgess and Wu, 2013), important for cognitive processes in general, inhibitory effects induced by the cathode might have prevented stronger effects of the anodal tDCS over left frontal and parietal sites. Other studies used a larger cathodal electrode (Rütsche et al., 2015; e.g., 10×10 cm in Hauser et al., 2013), or used an extracephalic return electrode (Pope et al., 2015), which might have mitigated or prevented similar disadvantageous effects.

A second limitation is that in this study a forced choice format (participants chose their answer from three options) was used. While this was similar to the work of Hauser et al. (2013) other groups like Rütsche et al. (2015) required the production of answers. This might be a reason why Rütsche and colleagues found detrimental effects of stimulation on the accuracy in easy problems, while this study did not. The task format used in the current study might have allowed the participants to reconsider their answer in light of the displayed options. However, the comparably high accuracy in small problems in this study (additions $M = 98.00\%$; subtractions $M = 96.76\%$) and in the study of Rütsche et al. (2015); stimulated $M = 97.82\%$; sham $M = 98.78\%$) speaks against this notion.

In conclusion, neither left frontal, nor left parietal stimulation led to a general improvement of arithmetic or working memory performance. However, there was a significant stimulation effect indicating an acceleration of training gains in large subtractions by left frontal stimulation. As stimulation effects on training and learning seem to be stronger than on performance *per se* (Simonsmeier et al., 2018), the effects might have been too small to enhance performance but still strong enough to improve procedural training. Hence, tDCS might be suited best to improve later performance when applied during learning or training while its potential to improve skills or their application in the sense of a sole performance enhancer remains ambiguous.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee at the University of Graz, University of Graz, Graz, Austria. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JM, CB, MN, and RG: study conception and design. JM and CB: data collection and analysis. JM, CB, MN, and RG: writing and/or critically revising the manuscript.

FUNDING

This study was funded by the Austrian Science Fund (FWF), project number: P30050-GBL.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00017/full#supplementary-material

REFERENCES

Aschenbrenner, S., Tucha, O., and Lange, K. W. (2000). *Regensburger Wortflüssigkeits-Test: RWT*. Göttingen: Hogrefe.

Borkenau, P., and Ostendorf, F. (2008). *NEO-Fünf-Faktoren-Inventar nach Costa und McCrae*. Göttingen: Hogrefe.

Brunoni, A. R., and Vanderhasselt, M.-A. (2014). Working memory improvement with non-invasive brain stimulation of the dorsolateral prefrontal cortex: a systematic review and meta-analysis. *Brain Cogn.* 86, 1–9. doi: 10.1016/J.BANDC.2014.01.008

Burgess, P. W., and Wu, H.-C. (2013). "Rostral prefrontal cortex (Broadmann area 10): metacognition in the brain," in *Principles of Frontal Lobe Function, 2nd Edn*, eds D. T. Stuss and R. T. Knight (New York, NY: OUT), 524–534.

Campbell, J. I. D., and Xue, Q. (2001). Cognitive arithmetic across cultures. *J. Exp. Psychol. Gen.* 130, 299–315. doi: 10.1037//0096-3445.130.2.299

Cavanagh, J. F., and Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends Cogn. Sci.* 18, 414–421. doi: 10.1016/J.TICS.2014.04.012

Clemens, B., Jung, S., Zvyagintsev, M., Domahs, F., and Willmes, K. (2013). Modulating arithmetic fact retrieval: a single-blind, sham-controlled tDCS study with repeated fMRI measurements. *Neuropsychologia* 51, 1279–1286. doi: 10.1016/j.neuropsychologia.2013.03.023

De Smedt, B., Grabner, R. H., and Studer, B. (2009). Oscillatory EEG correlates of arithmetic strategy use in addition and subtraction. *Exp. Brain Res.* 195, 635–642. doi: 10.1007/s00221-009-1839-9

Destefano, D., and LeFevre, J.-A. (2004). The role of working memory in mental arithmetic. *Eur. J. Cogn. Psychol.* 16, 353–386. doi: 10.1080/09541440244000328

Earle, J. B. B., Garcia-Dergay, P., Manniello, A., and Dowd, C. (1996). Mathematical cognitive style and arithmetic sign comprehension: a study of EEG alpha and theta activity. *Int. J. Psychophysiol.* 21, 1–13. doi: 10.1016/0167-8760(95)00039-9

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

Gill, J., Shah-Basak, P. P., and Hamilton, R. (2015). It's the thought that counts: examining the task-dependent effects of transcranial direct current stimulation on executive function. *Brain Stimul.* 8, 253–259. doi: 10.1016/J.BRS.2014.10.018

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., and Neuper, C. (2009). To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. *Neuropsychologia* 47, 604–608. doi: 10.1016/j.neuropsychologia.2008.10.013

Grabner, R. H., and De Smedt, B. (2011). Neurophysiological evidence for the validity of verbal strategy reports in mental arithmetic. *Biol. Psychol.* 87, 128–136. doi: 10.1016/j.biopsycho.2011.02.019

Grabner, R. H., and De Smedt, B. (2012). Oscillatory EEG correlates of arithmetic strategies: a training study. *Front. Psychol.* 3:428. doi: 10.3389/fpsyg.2012.00428

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2013). MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* 7:267. doi: 10.3389/fnins.2013.00267

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage* 86, 446–460. doi: 10.1016/j.neuroimage.2013.10.027

Hauser, T. U., Rotzer, S., Grabner, R. H., Mérillat, S., and Jäncke, L. (2013). Enhancing performance in numerical magnitude processing and mental arithmetic using transcranial Direct Current Stimulation (tDCS). *Front. Hum. Neurosci.* 7:244. doi: 10.3389/fnhum.2013.00244

Ishii, R., Canuet, L., Ishihara, T., Aoki, Y., Ikeda, S., Hata, M., et al. (2014). Frontal midline theta rhythm and gamma power changes during focused attention on mental calculation: an MEG beamformer analysis. *Front. Hum. Neurosci.* 8:406. doi: 10.3389/fnhum.2014.00406

Jasper, H. H. (1958). The ten-twenty electrode system of the international federation. *Electroencephalogr. Clin. Neurophysiol.* 10, 371–375.

Kasahara, K., Tanaka, S., Hanakawa, T., Senoo, A., and Honda, M. (2013). Lateralization of activity in the parietal cortex predicts the effectiveness of bilateral transcranial direct current stimulation on performance of a mental calculation task. *Neurosci. Lett.* 545, 86–90. doi: 10.1016/j.neulet.2013.04.022

Klein, E., Mann, A., Huber, S., Bloechle, J., Willmes, K., Karim, A. A., et al. (2013). Bilateral bi-cephalic tdcs with two active electrodes of the same polarity modulates bilateral cognitive processes differentially. *PLoS ONE* 8, 1–11. doi: 10.1371/journal.pone.0071607

Liepmann, D., Beauducel, A., Brocke, B., and Amthauer, R. (2007). *I-S-T 2000R Intelligenz-Struktur-Test 2000 R*. Göttingen: Hogrefe.

Menon, V. (2015). "Arithmetic in the child and adult brain," in *Handbook of Numerical Cognition*, eds R. C. Cohen-Kadosh and A. Dowker (Oxford: Oxford University Press), 502–530.

Missonnier, P., Deiber, M.-P., Gold, G., Millet, P., Gex-Fabry Pun, M., Fazio-Costa, L., et al. (2006). Frontal theta event-related synchronization: comparison of directed attention and working memory load effects. *J. Neural Transm.* 113, 1477–1486. doi: 10.1007/s00702-005-0443-9

Neisser, U. C., Boodoo, G., Bouchard, T. J. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: knowns and unknowns. *Am. Psychol.* 51, 77–101. doi: 10.1037/0003-066X.51.2.77

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., et al. (2008). Transcranial direct current stimulation: state of the art 2008. *Brain Stimul.* 1, 206–223. doi: 10.1016/j.brs.2008.06.004

OECD. (2016). *Skills Matter: Further Results From the Survey of Adult Skills, OECD Skills Studies*. Paris: OECD Publishing. doi: 10.1787/9789264258051-en

Parsons, S., and Bynner, J. (1997). Numeracy and employment. *Edu. Train.* 39, 43–51. doi: 10.1108/00400919710164125

Parsons, S., and Bynner, J. (2005). *Does Numeracy Matter More ? National Research and Development Centre for Adult Literacy and Numeracy.* London, UK: National Research and Development Centre for Adult Lliteracy and Numeracy.

Paulus, W. (2011). Transcranial electrical stimulation (tES - tDCS; tRNS, tACS) methods. *Neuropsychol. Rehabil.* 21, 602–617. doi: 10.1080/09602011.2011.557292

Pena-Gomez, C., Sala-Lonch, R., Junque, C., Clemente, I. C. I. C., Vidal, D., Baragllo, N., et al. (2012). Modulation of large-scale brain networks by transcranial direct current stimulation evidenced by resting-state functional MRI. *Brain Stimul.* 5, 252–263. doi: 10.1016/j.brs.2011.08.006

Polanía, R., Nitsche, M. A., and Paulus, W. (2011). Modulating functional connectivity patterns and topological functional organization of the human brain with transcranial direct current stimulation. *Hum. Brain Mapp.* 32, 1236–1249. doi: 10.1002/hbm.21104

Pope, P. A., Brenton, J. W., and Miall, R. C. (2015). Task-specific facilitation of cognition by anodal transcranial direct current stimulation of the prefrontal cortex. *Cerebral Cortex* 25, 4551–4558. doi: 10.1093/cercor/bhv094

Price, G. R., Mazzocco, M. M. M., and Ansari, D. (2013). Why mental arithmetic counts: brain activation during single digit arithmetic predicts high school math scores. *J. Neurosci.* 33, 156–163. doi: 10.1523/JNEUROSCI.2936-12.2013

Reynolds, C. R. (2002). *Comprehensive Trail Making Test (CTMT)*. Austin: Pro-Ed.

Romei, V., Thut, G., and Silvanto, J. (2016). Information-based approaches of noninvasive transcranial brain stimulation. *Trends Neurosci.* 39, 782–795. doi: 10.1016/J.TINS.2016.09.001

Rütsche, B., Hauser, T. U., Jäncke, L., and Grabner, R. H. (2015). When problem size matters: differential effects of brain stimulation on arithmetic problem solving and neural oscillations. *PLoS ONE* 10:e0120665. doi: 10.1371/journal.pone.0120665

Sarkar, A., Dowker, A., and Kadosh, R. C. (2014). Cognitive enhancement or cognitive cost: trait-specific outcomes of brain stimulation in the case of mathematics anxiety. *J. Neurosci.* 34, 16605–16610. doi: 10.1523/JNEUROSCI.3129-14.2014

Schroeder, P. A., Dresler, T., Bahnmueller, J., Artemenko, C., Cohen Kadosh, R., and Nuerk, H.-C. (2017). Cognitive enhancement of numerical and arithmetic capabilities: a mini-review of available transcranial electric stimulation studies. *J. Cogn. Enhanc.* 1, 39–47. doi: 10.1007/s41465-016-0006-z

Shalev, R. S., Auerbach, J., Manor, O., and Gross-Tsur, V. (2000). Developmental dyscalculia: prevalence and prognosis. *Europ. Child Adolesc. Psychiatry* 9, 58–64. doi: 10.1007/s007870070009

Silvanto, J., Muggleton, N., and Walsh, V. (2008). State-dependency in brain stimulation studies of perception and cognition. *Trends Cogn. Sci.* 12, 447–454. doi: 10.1016/J.TICS.2008.09.004

Simonsmeier, B. A., Grabner, R. H., Hein, J., Krenz, U., and Schneider, M. (2018). Electrical brain stimulation (tES) improves learning more than performance: a meta-analysis. *Neurosci. Biobehav. Rev.* 84, 171–181. doi: 10.1016/J.NEUBIOREV.2017.11.001

Soltanlou, M., Artemenko, C., Dresler, T., Fallgatter, A. J., Nuerk, H.-C., and Ehlis, A.-C. (2019). Oscillatory EEG changes during arithmetic learning in children. *Dev. Neuropsychol.* 44, 1–14. doi: 10.1080/87565641.2019.1586906

Steingrüber, H. J., and Lienert, G. A. (1971). *Hand-Dominanz-Test: HDT*. Göttingen: Hogrefe.

Zaehle, T., Sandmann, P., Thorne, J. D., Jäncke, L., and Herrmann, C. S. (2011). Transcranial direct current stimulation of the prefrontal cortex modulates working memory performance: combined behavioural and electrophysiological evidence. *BMC Neurosci.* 12:2. doi: 10.1186/1471-2202-12-2

Conflict of Interest: 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.

frontiers in Human Neuroscience

PERSPECTIVE
published: 20 March 2020
doi: 10.3389/fnhum.2020.00100

The Challenge of Modeling the Acquisition of Mathematical Concepts

Alberto Testolin [1,2]*

[1]*Department of General Psychology, University of Padova, Padova, Italy,* [2]*Department of Information Engineering, University of Padova, Padova, Italy*

OPEN ACCESS

Edited by:
Elise Klein,
University of Tübingen, Germany

Reviewed by:
Xinlin Zhou,
Beijing Normal University, China
Tom Verguts,
Ghent University, Belgium

***Correspondence:**
Alberto Testolin
alberto.testolin@unipd.it

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *13 November 2019*
Accepted: *04 March 2020*
Published: *20 March 2020*

Citation:
Testolin A (2020) The Challenge of Modeling the Acquisition of Mathematical Concepts. Front. Hum. Neurosci. 14:100. doi: 10.3389/fnhum.2020.00100

As a full-blown research topic, numerical cognition is investigated by a variety of disciplines including cognitive science, developmental and educational psychology, linguistics, anthropology and, more recently, biology and neuroscience. However, despite the great progress achieved by such a broad and diversified scientific inquiry, we are still lacking a comprehensive theory that could explain how numerical concepts are learned by the human brain. In this perspective, I argue that computer simulation should have a primary role in filling this gap because it allows identifying the finer-grained computational mechanisms underlying complex behavior and cognition. Modeling efforts will be most effective if carried out at cross-disciplinary intersections, as attested by the recent success in simulating human cognition using techniques developed in the fields of artificial intelligence and machine learning. In this respect, deep learning models have provided valuable insights into our most basic quantification abilities, showing how numerosity perception could emerge in multi-layered neural networks that learn the statistical structure of their visual environment. Nevertheless, this modeling approach has not yet scaled to more sophisticated cognitive skills that are foundational to higher-level mathematical thinking, such as those involving the use of symbolic numbers and arithmetic principles. I will discuss promising directions to push deep learning into this uncharted territory. If successful, such endeavor would allow simulating the acquisition of numerical concepts in its full complexity, guiding empirical investigation on the richest soil and possibly offering far-reaching implications for educational practice.

Keywords: computational modeling, artificial neural networks, deep learning, number sense, symbol grounding, mathematical learning, embodied cognition, material culture

INTRODUCTION

Despite the importance of mathematics in modern societies, the cognitive foundations of mathematical learning are still mysterious and hotly debated. At the one end of the bridge, the idealistic view conceives mathematical concepts as purely abstract entities that humans discover using logical reasoning; at the other end, empiricists argue that mathematics is the product of our sensory experiences, and therefore it is essentially an activity of construction (Brown, 2012). A somehow intermediate position is taken by modern neurocognitive theories, which identify a set of "core" brain systems specifically evolved to support basic intuitions about quantity (Butterworth, 1999; Feigenson et al., 2004; Piazza, 2010; Dehaene, 2011) but also acknowledge that higher-level numerical knowledge has materialized only recently, via cultural practices supported by language and symbolic reference (Núñez, 2017).

In recent years, the finding that measures of basic quantification skills correlate to later mathematical achievement (e.g., Halberda et al., 2008; Libertus et al., 2011; Starr et al., 2013) has led to the hypothesis that our "number sense" might indeed constitute the starting point to learn more complex mathematical concepts. However, the relationship between numerosity perception and symbolic math remains controversial (Negen and Sarnecka, 2015; Schneider et al., 2017; Wilkey and Ansari, 2019), calling for a deeper theoretical investigation that should be carried out with the support of formal models.

Here I will argue that the quest for artificial intelligence provides an extremely rich soil for the development of a computational theory of mathematical learning. Indeed, although computers largely outperform humans on numerical tasks requiring the mere application of syntactic manipulations (e.g., performing algebraic operations on large numbers, or iteratively computing the value of a function), they are completely blind about the *meaning* of such operations because they lack a conceptual semantics of number. Grounding abstract symbols into some form of intrinsic meaning is a longstanding issue in artificial intelligence (Searle, 1980; Harnad, 1990), and mathematics likely constitutes the most challenging domain for investigating how high-level knowledge could be linked to bottom-up, sensorimotor primitives (Leibovich and Ansari, 2016).

By framing a theory in computational terms, scientists are forced to adopt a precise, formal language, because all the details of the theory should be explicitly stated to simulate it on a computer. Modeling also requires to carefully think about the tasks that are being simulated and the possible ways in which a computational device can (or cannot) solve them. In this perspective article, I will focus in particular on *connectionist* models, where cognition is conceived as an emergent property of networks of units that self-organize according to physical principles (Rumelhart and McClelland, 1986; Elman et al., 1996; McClelland et al., 2010). According to this view, knowledge is implicitly stored in the connections among neurons, and learning processes adaptively change the strength of these connections according to experience. Notably, the recent breakthroughs in *deep learning* (LeCun et al., 2015) have revealed the true potential of this approach, by showing how machines endowed with domain-general learning mechanisms can simulate a variety of high-level cognitive skills, ranging from visual object recognition (He et al., 2016) to natural language understanding (Devlin et al., 2018) and strategic planning (Silver et al., 2017).

Computational Models of Basic Quantification Skills

According to the "number sense" view, numerical cognition is grounded in basic quantification skills, such as the ability to rapidly estimate the number of items in a visual display (Dehaene, 2011). Numerosity is thus conceived as a primary perceptual attribute (Anobile et al., 2016) processed by a specialized (and possibly innate) system yielding an approximate representation of numerical quantity (Feigenson et al., 2004). The seminal neural network model by Dehaene and Changeux (1993) incorporated these principles: numerosity perception was hardwired in the model, reflecting the assumption that this ability is present at birth. Successive models revisited this nativist stance, by showing that numerosity representations can emerge as a result of learning and sensory experience (Verguts and Fias, 2004). In particular, recent work based on unsupervised deep learning has demonstrated that human-like numerosity perception can emerge in multi-layer neural networks that learn a hierarchical generative model of the sensory data (Stoianov and Zorzi, 2012; Zorzi and Testolin, 2018; see **Figure 1A**).

Deep learning models account for a wide range of empirical phenomena in the number sense literature. They can accurately simulate Weber-like responses in numerosity comparison tasks (Stoianov and Zorzi, 2012), also accounting for congruency effects (Zorzi and Testolin, 2018) and for the fine-grained contribution of non-numerical magnitudes in biasing behavioral responses (Testolin et al., 2019). Notably, the number acuity of randomly initialized deep networks rivals that of newborns, and its gradual development follows trajectories similar to those observed in human longitudinal studies (Testolin et al., 2020). Deep networks have also been successfully tested in subitizing (Wever and Runia, 2019) and numerosity estimation tasks (Chen et al., 2018). Last, but not least, artificial neurons often reproduce neurophysiological properties observed in single-cell recording studies, for example by exhibiting number-sensitive tuning functions (Zorzi and Testolin, 2018; Nasr et al., 2019).

Several questions remain under investigation: Is it possible to fully disentangle numerosity from continuous magnitudes by only relying on unsupervised learning (Zanetti et al., 2019)? Can generative models generalize to unseen numerosities (Zhao et al., 2018)? Are there computational limitations in tracking multiple objects in dynamic scenes (Cenzato et al., 2019)? What is the contribution of explicit feedback and multi-sensory integration in shaping numerosity representations? How do deep learning models map into the cortical processing hierarchy? Nevertheless, despite these open questions, we can safely argue that deep learning has paved the way toward a computational theory about the origin of our number sense, confirming the appeal of deep networks as models of human sensory processing (Testolin and Zorzi, 2016; Yamins and DiCarlo, 2016; Testolin et al., 2017). Unfortunately, simulating the transition from approximate to symbolic numbers turns out to be much more challenging, as we discuss in the next section.

Modeling the Acquisition of Higher-Level Mathematical Concepts

One of the most ambitious questions to be addressed is whether deep learning models could develop even more sophisticated numerical abilities, such as those involving arithmetic and symbolic math. Symbolic reasoning is notoriously difficult for connectionist models (Marcus, 2003), and despite recent progress, deep neural networks still struggle with tasks requiring procedural and compositional knowledge (Garnelo and Shanahan, 2019).

Only a few modeling studies have investigated how arithmetic could be learned by artificial neural networks. Since early

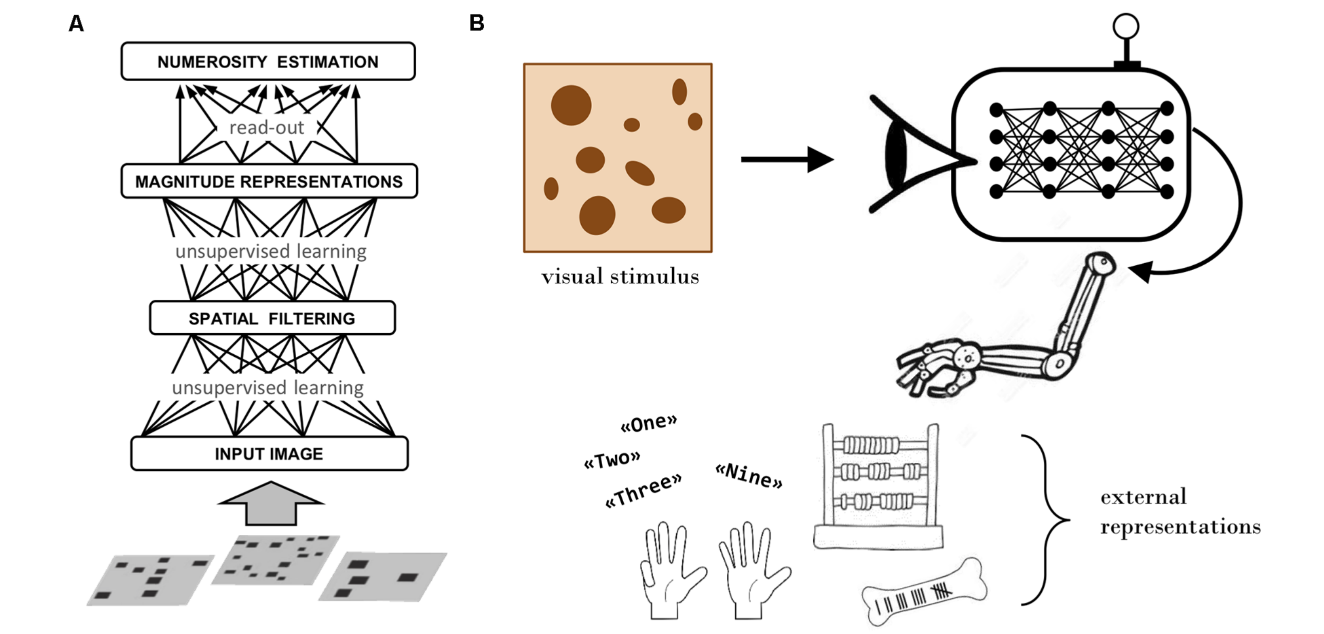


FIGURE 1 | Deep learning models. **(A)** Schematic representation of an unsupervised deep learning model that simulates human numerosity perception. Adapted from Zorzi and Testolin (2018). **(B)** Sketch of the proposed modeling framework, which extends the basic numerosity perception model (entirely confined within the agent's brain) by introducing the ability to interact with the external environment to create and manipulate material representations.

attempts, associative memories have been used to simulate mental calculation as a process of storage and retrieval of arithmetic facts (McCloskey and Lindemann, 1992): during the learning phase, the two arguments and the result of a simple operation (e.g., single-digit multiplication) are given as input to an associative memory, whose learning goal is to accurately store them as a global, stable state. During the testing phase, only the operands are given, and the network must recover the missing information (i.e., the result) by gradually settling into the correct configuration. Building on this approach, successive simulations have shown that numerosity-based ("semantic") representations can facilitate the learning of arithmetic facts (Zorzi et al., 2005) and equivalence problems (Mickey and McClelland, 2014). Others have shown that multi-digit addition and subtraction (but not multiplication) can be acquired through end-to-end supervised learning from pixel-level images (Hoshen and Peleg, 2015). One critical limitation of these approaches, however, is that they conceive arithmetic learning as a mere process of storing and recall, which gradually develops through the massive reiteration of all possible arithmetic facts that need to be learned. Besides being psychologically implausible and computationally unfeasible, this approach does not guarantee that the system will be able to generalize the acquired knowledge to unseen numbers and, even less, to exploit the acquired knowledge to more effectively learn new mathematical concepts.

The challenge of developing learning models that can exhibit algebraic generalization with the robustness and flexibility exhibited by humans is so fundamental that major players in deep learning research are intensively investigating these issues. For example, Google's DeepMind company has recently evaluated several deep learning models on a set of benchmark problems taken from UK national school mathematics curriculums, covering arithmetic, algebra, elementary calculus, et cetera (Saxton et al., 2019). DeepMind's best model correctly solved only 14 out of 40 problems, which would be equivalent to an "E" grade. Although such difficulties have led some researchers to argue that neural networks are incapable of exhibiting compositional abilities (Marcus, 2018), others argue for the opposite (Baroni, 2020; Martin and Baggio, 2020).

Even the acquisition of the concept of *exact number* is still out of reach for deep networks, which often cannot generalize outside of the range of numerical values encountered during training (Trask et al., 2018). Integer numbers are one of the pillars of arithmetic, so they constitute the perfect testbed for developing and testing computational models of mathematical learning. Developmental studies show that integers are gradually acquired by children during formal education through the acquisition of number words and counting skills: Indeed, although sequential (item-by-item) enumeration skills are present in animal species (Platt and Johnson, 1971; Beran and Beran, 2004; Dacke and Srinivasan, 2008), even in humans counting is not culturally universal (Gordon, 2004) and there is evidence that young children and people from cultures lacking number words have an incomplete understanding of what it means for two sets of items to have exactly the same number of items (Izard et al., 2008, 2014).

Some authors have sought to characterize the acquisition of exact numbers as the semantic induction of a "cardinality principle" (Sarnecka and Carey, 2008). This hypothesis has been exemplified in a computational model based on Bayesian inference, which simulated the stage-like development of counting abilities by relying on a pre-determined set of "core" cognitive operations (Piantadosi et al., 2012). The repertoire of innate abilities included the capacity to exactly identify cardinalities up to 3, perform basic operations on sets (e.g., difference, union, intersection), retrieve the next or previous word from an ordered counting list, and to operate these functions recursively. Although such modeling approach offers a rational interpretation of the process that might underly the acquisition of an abstract cardinality principle, it assumes a certain amount of *a priori* symbolic knowledge and procedural skills, which is in contrast to empirical data suggesting, for example, that a complete understanding of the successor principle arises only after considerable interaction with the teaching environment (Davidson et al., 2012).

TOWARD A COMPREHENSIVE NEUROCOMPUTATIONAL FRAMEWORK

The Downplayed Role of External Representations

A central tenet of connectionist models is that semantics intrinsically emerges in a system interacting with its surrounding environment. However, this idea is usually superficially implemented in deep learning models, because the interaction is often limited to *passive* observation of statistical properties of the world (Zorzi et al., 2013). Taking inspiration from constructivist theories in developmental psychology, here I argue that a step forward will require to build computational models that learn by *actively* manipulating the environment, that is, by causally interacting with objects in their perceptual space. Crucially, the notion of "environment" should include embodiment (Lakoff and Núñez, 2000) and—most importantly—the social, cultural and educational environment (Vygotsky, 1980; Clark, 2011). Indeed, according to the Vygotskyan perspective, students actively construct abstract knowledge through interactions with teachers and peers, gradually moving their dependency on explicit forms of mediation to more implicit (internalized) forms (Walshaw, 2017).

The possibility to manipulate the environment greatly increases the complexity of the learning agent but also enables the functional use of external entities to create powerful representational systems, which can be manipulated in simple ways to get answers to difficult problems. The underlying assumption is that cultural evolution and history are foundational forces for the emergence of superior cognitive functions and that great intellectual achievements (such as the invention of mathematics) have been triggered by our ability to create artifacts serving as physical representations of abstract concepts. Some investigators have recently emphasized the role of material culture in numerical cognition (Menary, 2015; Overmann, 2016, 2018), for example by highlighting that our mental organization of numbers into an ordered "number line" might be related to the linearity of the material forms used to represent and manipulate them (Núñez, 2011). Primitive devices used for representing numbers date back to notched bones in the Paleolithic period (d'Errico et al., 2018) and clay tokens in the Neolithic period (Schmandt-Besserat, 1992), which predated the subsequent diffusion of abaci, positional systems and increasingly more sophisticated numerical notations (Menninger, 1992). However, despite the concept of external representations was foreseen in early connectionist theories[1], it has been seldomly explored in practice.

Learning to Create and Manipulate Symbolic Representations

We can now sketch a concrete proposal for building more realistic simulations of mathematical learning. The computational framework should incorporate the following key components, summarized in **Figure 1B**.

- **Perceptual system**. This is where computational modeling has been mostly focused (and successful) up to now (see Section "Computational Models of Basic Quantification Skills"). The challenge will be to scale-up the existing models to more realistic sensory input (e.g., naturalistic visual scenes) and to incorporate a larger repertoire of pattern recognition abilities, which should not only allow to approximately represent visual quantities but also to recognize structured configurations of object arrays (e.g., sequences of tally marks, geometric displacements of items, patterns encoded in an abacus, etc.) and symbolic notations (e.g., written digits and operands).
- **Embodiment**. Of particular interest to the development of exact numbers is finger counting (Butterworth, 1999; Andres et al., 2007; Domahs et al., 2012), which not only helps children to keep track and coordinate the production of number words (Alibali and DiRusso, 1999) but may also allow to organize numbers spatially (Fischer, 2008). Hand-based representations are ubiquitous across cultures (Bender and Beller, 2012) and play a key role in the subsequent acquisition of number words (Gunderson et al., 2015; Gibson et al., 2019), possibly influencing symbolic number processing even in adulthood (Domahs et al., 2010). It has been recently shown that neural networks can learn to count the number of items in visual displays and that the ability to sequentially point to individual objects helps in speeding up counting acquisition (Fang et al., 2018). A further step is taken by cognitive developmental robotics, which explores the instantiation of these principles in physically embodied agents (Di Nuovo and Jay, 2019). Interestingly, pointing gestures significantly improved counting accuracy in a humanoid robot, and learning was more effective when both fingers and words were provided as input (Rucinski et al., 2012; De La Cruz et al., 2014).
- **Material representations**. The ability to manipulate external objects might be the key missing piece for simulating the acquisition of exact numbers. Indeed, although hand gestures might serve as placeholders to

[1] See for example, the section "External Representations and Formal Reasoning" in Rumelhart et al. (1986).

learn more efficient arithmetic strategies (Siegler and Jenkins, 1989; for a computational account see Hansen et al., 2014), material representations allow for a much more precise encoding of numerical information. For example, the agent can learn to establish the cardinality of a set by organizing items in regular configurations that promote "groupitizing" (Starkey and McCandliss, 2014), or to exactly compare the cardinality of two sets by disposing of items in one-to-one correspondence. More sophisticated devices such as abaci and Cuisenaire rods further extend our ability to represent exact numbers, for example by exploiting inter-exponential relations to precisely (but compactly) encode large numbers, or to explicitly represent compositionality to promote generalization (Overmann, 2018).

- **Diversified learning signals**. In addition to unsupervised learning, the agent should exploit *reinforcement learning* (Sutton and Barto, 1998) to predict the outcome of its actions. This learning modality would also play a key role in simulating curiosity-driven behavior and active engagement with material representations. Notably, deep reinforcement learning has recently achieved impressive performance in difficult cognitive tasks, for example by discovering complex strategies in board games (Silver et al., 2017). However, learning through reinforcement can be challenging in the presence of very large action spaces (i.e., the correct action has to be chosen from a wide range of possible actions) and sparse rewards (i.e., feedback is given only once the whole task has been carried out). Taking inspiration from the notions of *transfer learning* and *curriculum learning* used in machine learning (Bengio et al., 2009) and from *shaping* procedures used in animal conditioning (Skinner, 1953), these issues can be mitigated by decomposing the task into simpler sub-tasks. For example, rather than rewarding only the trials where the agent has correctly counted all items in a display, rewards can be initially given every time the agent touches an object, to first promote the acquisition of sequential pointing skills. Similarly, the agent could first be rewarded simply for being able to accurately reproduce the abacus configuration corresponding to a specific number, rather than for being able to correctly manipulate the abacus to solve an addition problem. This idea of "gradually walking the agent through the word" also implies the exploitation of *supervised learning*, because explicit teaching signals must be used to stimulate learning by imitation and adult guidance.
- **Linguistic input**. Despite language might not be crucial for the acquisition of elementary numerical concepts (Gelman and Butterworth, 2005; Butterworth et al., 2008), it provides useful cues during the development of basic algebraic notions: for example, morphological cues allow single/plural distinction, number words can act as stable placeholders during counting acquisition, and learning natural language quantifiers seems a key step for mastering the ordering principle (Le Corre, 2014). A recent deep learning model has shown that learning quantifiers allows to more easily carry out approximate numerosity judgments (Pezzelle et al., 2018); however, the role of linguistic input for simulating the acquisition of exact numbers has yet to be explored. Furthermore, later in development language becomes the primary medium to acquire higher-level mathematical knowledge, hence it will need to be taken into account to design computational models approaching that level of complexity.

DISCUSSION

Symbolic numbers are a hallmark of human intelligence, but we are still lacking a comprehensive theory explaining how the brain learns to master them. Here I argued that computational modeling should have a primary role in this enterprise. Taking the acquisition of natural numbers as a case study, I emphasized the role of material representations in supporting the transition from approximate to symbolic numerical concepts. According to this view, exact numbers do not emerge from the mere association between number words and perceptual magnitudes: such mapping is strongly mediated by the acquisition of procedural skills (e.g., finger counting) and the ability to effectively manipulate representational devices (Leibovich and Ansari, 2016; Overmann, 2018; Carey and Barner, 2019).

In line with the idea that improved problem representation is a key mechanism for the joint development of conceptual and procedural knowledge (Rittle-Johnson et al., 2001), cognitive development in artificial agents must thus be supported by an adequate learning environment, which should provide feedback, teaching signals, and representational media commensurate with the current level of development. Notably, once a procedural skill has been mastered it might become internalized: the agent can simply "imagine" carrying out operations on the material device, without the need to physically operate over it. Some representations might thus serve just as intermediate steps for the acquisition of more abstract and efficient notations: as finger counting allows us to gradually grasp the meaning of number words, manipulating an abacus allows to ground numerical symbols into concrete visuospatial representations. A historical case that illustrates this perspective is the famous dispute between "abacists" and "algorists", which was undoubtedly won by the latter, who demonstrated the superiority of symbolic notation for carrying out arithmetic operations (see **Figure 2**). However, one might wonder whether Boethius could have mastered arithmetic algorithms without first grounding his numerical concepts into a set of more concrete representations.

In addition to providing a useful framework to interpret empirical findings, the proposed approach can raise important questions that would stimulate further theoretical and experimental work. For example, a critical aspect of our school system is to teach how to effectively discover useful strategies and representational schemes for solving difficult problems. In computational simulations, the necessity for appropriate teacher guidance stems from the fact that it is very difficult to invent new representations for problems we might wish to solve: it may even be that the process of

FIGURE 2 | Allegory of Arithmetic. Engraving from the encyclopedic book *Margarita Philosophica* by Gregor Reisch (1503) depicting the "abacists vs. algorists" debate. Arithmetica (female figure) is supervising a calculation contest between Pythagoras (right), represented as using a counting board, and Boethius (left), who embraces algorithmic calculation with Arabic numbers. The struggle of Pythagoras suggests who is going to be the winner. Reproduced from Wikipedia.

inventing such representations is one of our highest intellectual abilities (Rumelhart et al., 1986). Computational frameworks that allow simulating a more complex interaction between artificial agents and their learning environment might thus eventually provide insights also about the teaching practices that could be most effective to guide numerical development in our children.

AUTHOR CONTRIBUTIONS

AT is fully responsible for the content of this article.

FUNDING

This work was supported by the STARS grant "DEEPMATH" to AT from the University of Padova (Università degli Studi di Padova).

ACKNOWLEDGMENTS

I am grateful to Jay McClelland for the useful discussions that stimulated many ideas elaborated in this article.

REFERENCES

Alibali, M. W., and DiRusso, A. A. (1999). The function of gesture in learning to count: more than keeping track. *Cogn. Dev.* 56, 37–56. doi: 10.1016/s0885-2014(99)80017-3

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

Anobile, G., Cicchini, G. M., and Burr, D. C. (2016). Number as a primary perceptual attribute: a review. *Perception* 45, 5–31. doi: 10.1177/0301006615602599

Baroni, M. (2020). Linguistic generalization and compositionality in modern artificial neural networks. *Philos. Trans. R. Soc. B Biol. Sci.* 375:20190307. doi: 10.1098/rstb.2019.0307

Bender, A., and Beller, S. (2012). Nature and culture of finger counting: diversity and representational effects of an embodied cognitive tool. *Cognition* 124, 156–182. doi: 10.1016/j.cognition.2012.05.005

Bengio, Y., Louradour, J., Collobert, R., and Weston, J. (2009). "Curriculum learning," in *Proceedings of the 26th Annual International Conference on Machine Learning* (New York, NY, USA: ACM), 1–8.

Beran, M. J., and Beran, M. M. (2004). Chimpanzees remember the results of one-by-one addition of food items to sets over extended time periods. *Psychol. Sci.* 15, 94–99. doi: 10.1111/j.0963-7214.2004.01502004.x

Brown, J. R. (2012). *Platonism, Naturalism and Mathematical Knowledge.* New York, NY: Routledge.

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

Butterworth, B., Reeve, R., Reynolds, F., and Lloyd, D. (2008). Numerical thought with and without words: evidence from indigenous Australian children. *Proc. Natl. Acad. Sci. U S A* 105, 13179–13184. doi: 10.1073/pnas.0806045105

Carey, S., and Barner, D. (2019). Ontogenetic origins of human integer representations. *Trends Cogn. Sci.* 23, 823–835. doi: 10.1016/j.tics.2019.07.004

Cenzato, A., Testolin, A., and Zorzi, M. (2019). On the difficulty of learning and predicting the long-term dynamics of bouncing objects. *arXiv:1907.13494* [Preprint].

Chen, S. Y., Zhou, Z., Fang, M., and McClelland, J. L. (2018). "Can generic neural networks estimate numerosity like humans?," in *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, eds T. T. Rogers, M. Rau, X. Zhu and C. W. Kalish (Austin, TX: Cognitive Science Society), 202–207.

Clark, A. (2011). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension.* New York, NY: Oxford University Press.

d'Errico, F., Doyon, L., Colagé, I., Queffelec, A., Le Vraux, E., Giacobini, G., et al. (2018). From number sense to number symbols. An archaeological perspective. *Philos. Trans. R. Soc. B Biol. Sci.* 373:20160518. doi: 10.1098/rstb.2016.0518

Dacke, M., and Srinivasan, M. V. (2008). Evidence for counting in insects. *Anim. Cogn.* 11, 683–689. doi: 10.1007/s10071-008-0159-y

Davidson, K., Eng, K., and Barner, D. (2012). Does learning to count involve a semantic induction? *Cognition* 123, 162–173. doi: 10.1016/j.cognition.2011.12.013

De La Cruz, V. M., Di Nuovo, A., Di Nuovo, S., and Cangelosi, A. (2014). Making fingers and words count in a cognitive robot. *Front. Behav. Neurosci.* 8:13. doi: 10.3389/fnbeh.2014.00013

Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics.* New York, NY: Oxford University Press.

Dehaene, S., and Changeux, J.-P. J. (1993). Development of elementary numerical abilities: a neuronal model. *J. Cogn. Neurosci.* 5, 390–407. doi: 10.1162/jocn.1993.5.4.390

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2018). BERT: pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805* [Preprint].

Di Nuovo, A., and Jay, T. (2019). Development of numerical cognition in children and artificial systems: a review of the current knowledge and proposals for multi-disciplinary research. *Cogn. Comput. Syst.* 1, 2–11. doi: 10.1049/ccs.2018.0004

Domahs, F., Kaufmann, L., and Fischer, M. H. (2012). *Handy Numbers: Finger Counting and Numerical Cognition.* Lausanne: Special Research Topic in Frontiers Media SA.

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

Elman, J. L., Bates, E., Johnson, M., Karmiloff-smith, A., Parisi, D., and Plunkett, K. (1996). *Rethinking Innateness: A Connectionist Perspective on Development.* Cambridge, MA: MIT Press.

Fang, M., Zhou, Z., Chen, S. Y., and McClelland, J. L. (2018). "Can a recurrent neural network learn to count things?" in *Proceedings of the 40th Annual Conference of the Cognitive Science Society*, eds T. T. Rogers, M. Rau, X. Zhu, and C. W. Kalish (Austin, TX: Cognitive Science Society), 360–365.

Feigenson, L., Dehaene, S., and Spelke, E. S. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307–314. doi: 10.1016/j.tics.2004.05.002

Fischer, M. H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex* 44, 386–392. doi: 10.1016/j.cortex.2007.08.004

Garnelo, M., and Shanahan, M. (2019). Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. *Curr. Opin. Behav. Sci.* 29, 17–23. doi: 10.1016/j.cobeha.2018.12.010

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

Gibson, D. J., Gunderson, E. A., Spaepen, E., Levine, S. C., and Goldin-Meadow, S. (2019). Number gestures predict learning of number words. *Dev. Sci.* 22:e12791. doi: 10.1111/desc.12791

Gordon, P. (2004). Numerical cognition without words: evidence from amazonia. *Science* 306, 496–499. doi: 10.1126/science.1094492

Gunderson, E. A., Spaepen, E., Gibson, D., Goldin-Meadow, S., and Levine, S. C. (2015). Gesture as a window onto children's number knowledge. *Cognition* 144, 14–28. doi: 10.1016/j.cognition.2015.07.008

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

Hansen, S. S., McKenzie, C., and McClelland, J. L. (2014). "Two plus three is five: discovering efficient addition strategies without metacognition," in *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (Montreal, QC: Cognitive Science Society), 583–588.

Harnad, S. (1990). The symbol grounding problem. *Phys. D Nonlinar Phenom.* 42, 335–346. doi: 10.1016/0167-2789(90)90087-6

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Las Vegas, NV: IEEE), 770–778.

Hoshen, Y., and Peleg, S. (2015). "Visual learning of arithmetic operations," in *Proceedings of the 30th AAAI Conference on Artificial Intelligence*, (Phoenix, AZ: AAAI), 3733–3739.

Izard, V., Pica, P., Spelke, E. S., and Dehaene, S. (2008). Exact equality and successor function: two key concepts on the path towards understanding exact numbers. *Philos. Psychol.* 21, 491–505. doi: 10.1080/09515080802285354

Izard, V., Streri, A., and Spelke, E. S. (2014). Toward exact number: young children use one-to-one correspondence to measure set identity but not numerical equality. *Cogn. Psychol.* 72, 27–53. doi: 10.1016/j.cogpsych.2014.01.004

Lakoff, G., and Núñez, R. E. (2000). *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics Into Being.* New York, NY: Basic Books.

Le Corre, M. (2014). Children acquire the later-greater principle after the cardinal principle. *Br. J. Dev. Psychol.* 32, 163–177. doi: 10.1111/bjdp.12029

LeCun, Y., Bengio, Y., and Hinton, G. E. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Leibovich, T., and Ansari, D. (2016). The symbol-grounding problem in numerical cognition: a review of theory, evidence, and outstanding questions. *Can. J. Exp. Psychol.* 70, 12–23. doi: 10.1037/cep0000070

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

Marcus, G. F. (2003). *The Algebraic Mind: Integrating Connectionism and Cognitive Science.* Cambridge, MA: MIT Press.

Marcus, G. F. (2018). Deep learning: a critical appraisal. *arXiv:1801.00631* [Preprint].

Martin, A. E., and Baggio, G. (2020). Modelling meaning composition from formalism to mechanism. *Philos. Trans. R. Soc. B Biol. Sci.* 375:20190298. doi: 10.1098/rstb.2019.0298

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., et al. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends Cogn. Sci.* 14, 348–356. doi: 10.1016/j.tics.2010.06.002

McCloskey, M., and Lindemann, A. M. (1992). Mathnet: preliminary results from a distributed model of arithmetic fact retrieval. *Adv. Psychol.* 91, 365–409. doi: 10.1016/s0166-4115(08)60892-4

Menary, R. (2015). Mathematical cognition—a case of enculturation. *Open Mind* 25, 1–20. doi: 10.15502/9783958570818

Menninger, K. (1992). *Number Words and Number Symbols: A Cultural History of Numbers.* New York: Dover Pubns.

Mickey, K. W., and McClelland, J. L. (2014). "A neural network model of learning mathematical equivalence," in *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, (QC, Canada: Cognitive Science Society), 1012–1017.

Nasr, K., Viswanathan, P., and Nieder, A. (2019). Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. *Sci. Adv.* 5:eaav7903. doi: 10.1126/sciadv.aav7903

Negen, J., and Sarnecka, B. W. (2015). Is there really a link between exact-number knowledge and approximate number system acuity in young children? *Br. J. Dev. Psychol.* 33, 92–105. doi: 10.1111/bjdp.12071

Núñez, R. E. (2011). No innate number line in the human brain. *J. Cross. Cult. Psychol.* 42, 651–668. doi: 10.1177/0022022111406097

Núñez, R. E. (2017). Is there really an evolved capacity for number? *Trends Cogn. Sci.* 21, 409–424. doi: 10.1016/j.tics.2017.03.005

Overmann, K. A. (2016). The role of materiality in numerical cognition. *Quat. Int.* 405, 42–51. doi: 10.1016/j.quaint.2015.05.026

Overmann, K. A. (2018). Constructing a concept of number. *J. Numer. Cogn.* 4, 464–493. doi: 10.5964/jnc.v4i2.161

Pezzelle, S., Sorodoc, I.-T., and Bernardi, R. (2018). "Comparatives, quantifiers, proportions: a multi-task model for the learning of quantities from vision," in *Proceedings of the 2018 Conference of the Association for Computational Linguistics*, (New Orleans, LA: Association for Computational Linguistics), 419–430.

Piantadosi, S. T., Tenenbaum, J. B., and Goodman, N. D. (2012). Bootstrapping in a language of thought: a formal model of numerical concept learning. *Cognition* 123, 199–217. doi: 10.1016/j.cognition.2011.11.005

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

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

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

Rucinski, M., Cangelosi, A., and Belpaeme, T. (2012). "Robotic model of the contribution of gesture to learning to count," in *IEEE International Conference on Development and Learning and Epigenetic Robotics* (San Diego, CA, USA: IEEE), 1–6. doi: 10.1109/DevLrn.2012.6400579

Rumelhart, D. E., and McClelland, J. L. (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations*. Cambridge, MA: MIT Press.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., and Hinton, G. E. (1986). "Schemata and sequential thought processes in PDP models," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition Volume 2: Psychological and Biological Models* eds J. L. McClelland and D. E. Rumelhart (Cambridge, MA: MIT Press), 7–57.

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

Saxton, D., Grefenstette, E., Hill, F., and Kohli, P. (2019). "Analysing mathematical reasoning abilities of neural models," in *International Conference on Learning Representations*, (New Orleans, LA: ICLR), 1–17.

Schmandt-Besserat, D. (1992). *Before Writing: From Counting to Cuneiform.* Austin, TX: University of Texas Press.

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372

Searle, J. R. (1980). Minds, brains, and programs. *Behav. Brain Sci.* 3, 417–424. doi: 10.1017/S0140525X00005756

Siegler, R. S., and Jenkins, E. (1989). *How Children Discover New Strategies.* New York, NY: Psychology Press.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., et al. (2017). Mastering the game of Go without human knowledge. *Nature* 550, 354–359. doi: 10.1038/nature24270

Skinner, B. (1953). *Science and Human Behavior.* New York, NY: MacMillan.

Starkey, G. S., and McCandliss, B. D. (2014). The emergence of "groupitizing" in children's numerical cognition. *J. Exp. Child Psychol.* 126, 120–137. doi: 10.1016/j.jecp.2014.03.006

Starr, A., Libertus, M. E., and Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proc. Natl. Acad. Sci. U S A* 110, 18116–18120. doi: 10.1073/pnas.1302751110

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

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning.* Cambridge, MA: MIT Press.

Testolin, A., Dolfi, S., Rochus, M., and Zorzi, M. (2019). Perception of visual numerosity in humans and machines *arXiv:1907.06996* [Preprint].

Testolin, A., Stoianov, I., and Zorzi, M. (2017). Letter perception emerges from unsupervised deep learning and recycling of natural image features. *Nat. Hum. Behav.* 1, 657–664. doi: 10.1038/s41562-017-0186-2

Testolin, A., and Zorzi, M. (2016). Probabilistic models and generative neural networks: towards an unified framework for modeling normal and impaired neurocognitive functions. *Front. Comput. Neurosci.* 10:73. doi: 10.3389/fncom.2016.00073

Testolin, A., Zou, W. Y., and McClelland, J. L. (2020). Numerosity discrimination in deep neural networks: initial competence, developmental refinement and experience statistics. *Dev. Sci.* doi: 10.1111/desc.12940 [Epub ahead of print].

Trask, A., Hill, F., Reed, S., Rae, J., Dyer, C., and Blunsom, P. (2018). Neural arithmetic logic units. *arXiv:1808.00508* [Preprint].

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

Vygotsky, L. S. (1980). *Mind in Society: The Development of Higher Psychological Processes.* Cambridge, MA: Harvard University Press.

Walshaw, M. (2017). Understanding mathematical development through Vygotsky. *Res. Math. Educ.* 19, 293–309. doi: 10.1080/14794802.2017.1379728

Wever, R., and Runia, T. F. H. (2019). Subitizing with variational autoencoders. *arXiv:1808.00257* [Preprint].

Wilkey, E. D., and Ansari, D. (2019). Challenging the neurobiological link between number sense and symbolic numerical abilities. *Ann. N Y Acad. Sci.* doi: 10.1111/nyas.14225 [Epub ahead of print].

Yamins, D. L. K., and DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nat. Neurosci.* 19, 356–365. doi: 10.1038/nn.4244

Zanetti, A., Testolin, A., Zorzi, M., and Wawrzynski, P. (2019). Numerosity representation in InfoGAN: an empirical study *Advances in Computational Intelligence. IWANN*, eds I. Rojas, G. Joya and A. Catala (Cham: Springer International Publishing), 49–60.

Zhao, S., Ren, H., Yuan, A., Song, J., Goodman, N., and Ermon, S. (2018). "Bias and generalization in deep generative models: an empirical study," in *32nd*

International Conference on Neural Information Processing Systems (Montréal, QC: NeurIPS), 10792–10801.

Zorzi, M., Stoianov, I., and Umiltà, C. (2005). "Computational modeling of numerical cognition," in *Handbook of Mathematical Cognition*, ed. J. Campbell (New York, NY: Psychology Press), 67–84.

Zorzi, M., and Testolin, A. (2018). An emergentist perspective on the origin of number sense. *Philos. Trans. R. Soc. B Biol. Sci.* 373:20170043. doi: 10.1098/rstb.2017.0043

Zorzi, M., Testolin, A., and Stoianov, I. P. I. P. (2013). Modeling language and cognition with deep unsupervised learning: a tutorial overview. *Front. Psychol.* 4:515. doi: 10.3389/fpsyg.2013.00515

Conflict of Interest: 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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 24 March 2020
doi: 10.3389/fpsyg.2020.00529

A Finger-Based Numerical Training Failed to Improve Arithmetic Skills in Kindergarten Children Beyond Effects of an Active Non-numerical Control Training

Ulrike Schild*, Anne Bauch and Hans-Christoph Nuerk

Department of Psychology, University of Tübingen, Tübingen, Germany

OPEN ACCESS

Edited by:
Reuven Babai,
Tel Aviv University, Israel

Reviewed by:
Francesco Sella,
University of Sheffield,
United Kingdom
Catherine Thevenot,
Université de Lausanne, Switzerland

****Correspondence:***
Ulrike Schild
ulrike.schild@uni-tuebingen.de

Specialty section:
This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: *11 November 2019*
Accepted: *05 March 2020*
Published: *24 March 2020*

Citation:
Schild U, Bauch A and Nuerk H-C (2020) A Finger-Based Numerical Training Failed to Improve Arithmetic Skills in Kindergarten Children Beyond Effects of an Active Non-numerical Control Training.
Front. Psychol. 11:529.
doi: 10.3389/fpsyg.2020.00529

It is widely accepted that finger and number representations are associated: many correlations (including longitudinal ones) between finger gnosis/counting and numerical/arithmetical abilities have been reported. However, such correlations do not necessarily imply causal influence of early finger-number training; even in longitudinal designs, mediating variables may be underlying such correlations. Therefore, we investigated whether there may be a causal relation by means of an extensive experimental intervention in which the impact of finger-number training on initial arithmetic skills was tested in kindergarteners to see whether they benefit from the intervention even before they start formal schooling. The experimental group received 50 training sessions altogether for 10 weeks on a daily basis. A control group received phonology training of a similar duration and intensity. All children improved in the arithmetic tasks. To our surprise and contrary to most accounts in the literature, the improvement shown by the experimental training group was not superior to that of the active control group. We discuss conceptual and methodological reasons why the finger-number training employed in this study did not increase the initial arithmetic skills beyond the unspecific effects of the control intervention.

Keywords: finger-number associations, initial arithmetic skills, embodiment, intervention, children

INTRODUCTION

Being able to competently deal with numbers is a fundamental skill in our society. Recently, the interest of researchers has turned to precursor abilities of mathematical achievement like approximate number processing (for a review see De Smedt et al., 2013; Libertus et al., 2016), spatial skills (e.g., Cipora et al., 2015), spatial number associations (e.g., Cipora et al., 2015), verbal number skills (e.g., Libertus et al., 2016), counting (e.g., Nguyen et al., 2016), mathematical language (Purpura et al., 2017) or base-10-knowledge (Moeller et al., 2011). Another of these potential precursors might be finger representation or finger gnosis (see Moeller and Nuerk, 2012 for a discussion). In turn, finger gnosis may serve to build up associations between fingers and numbers. It has been argued that finger representations might be another important precursor for initial arithmetic skills as they provide the child with an embodied representation of numbers developmentally located at the transition between early non-verbal representations and cultural symbolic representations. Such arguments rest on theoretical considerations (e.g., Moeller et al., 2011; Moeller and Nuerk, 2012) and observed correlations; however, whether earlier finger-number relations really have effects on later arithmetic skills has rarely been investigated. Therefore, the core purpose of this study was to examine intervention effects of finger-number associations on early arithmetic skills.

There is solid evidence now that finger and number representation are associated. First evidence was provided by Gerstmann (1940) who described neurological syndromes like finger agnosia, agraphia, acalculia and a disorientation for right and left that occurred together. This combination of deficits suggests that the same brain regions are responsible for the underlying processes. Over the last decades, studies using brain imaging techniques supported this close connection. Overlapping brain regions were found for finger representations and brain areas involved in number counting (e.g., Tschentscher et al., 2012) or arithmetic calculations (e.g., Berteletti and Booth, 2015). Many behavioral studies in adults also support an association of, for example, finger representation and counting [but see Brozzoli et al. (2008) for a dominance of a mental-number line representation when directly contrasted with finger-number representations], of finger representation and cardinality, and of finger representation and arithmetic (for a short overview see Di Luca and Pesenti, 2011). However for behavioral, as well as for brain imaging studies, most evidence so far is correlational – a truly *causal* relation between finger representation and numerical/arithmetic skills by manipulating finger knowledge and built-up representations has rarely been shown. Whether children refine their finger representations in parallel or in mutual interaction with the acquisition of their initial numerical skills or whether a good finger representation is beneficial or even necessary for developing numerical representations and/or numerical competencies is an open and controversial question in numerical development and education (Moeller and Nuerk, 2012).

A growing number of studies showed that finger representation (or finger gnosis) is associated with basic numerical skills (Costa et al., 2011) and that finger gnosis can predict later numerical skills (Fayol et al., 1998; Noël, 2005). However, the explained variance tends to be small. This was particularly the case when possible third variables like general cognitive ability were taken into account, and a sufficient number of participants was tested (Penner-Wilger et al., 2007, 2009; Kohn et al., 2015; Poltz et al., 2015; Wyschkon et al., 2015; Long et al., 2016; Wasner et al., 2016). Nonetheless, finger representations do seem to affect numerical processing in both children and adults as shown, for example, in the finger-based sub-base five effects (e.g., Domahs et al., 2008, 2010). To additionally investigate the role of finger gnosis as a precursor for later arithmetic skills, a sub-purpose of our study was to look at the predictive value of finger gnosis at pre-intervention for initial mathematical skills at post-intervention.

However, it is important to distinguish between finger gnosis or finger representations, finger-number associations and direct finger use in finger counting and arithmetic tasks.

Concerning finger use in number tasks, when children start to communicate about numbers or when they learn to count, they often use their fingers (e.g., showing their age with their fingers). This is even true for blind children (Crollen et al., 2011a; but see Crollen et al., 2014 for the role of visual experience in finger-number associations) or for children without hands who use their phantom fingers to count (Poeck, 1964). Even later when starting to acquire addition and subtraction skills many children use their fingers (e.g., Butterworth, 1999). Furthermore, when prevented from using their fingers by interfering hand movements arithmetic performance seems to drop (Crollen and Noël, 2015). This shows that fingers are used in a numerical and arithmetic context but does not imply that this finger-number association leads to better arithmetic performance.

Children who use their fingers directly might have ‘good’ finger representations and finger-number associations. In contrast, children who do not use their fingers directly, might have either ‘poor’ finger representation and in turn ‘poor’ finger-number associations, which prevents them from using their fingers. Or they might have ‘very good and stable’ finger representations and finger-number associations, but are no longer in need of using their fingers directly, because they have already built up good abstract numerical representations. Thus, conclusions about the relation between direct finger use and underlying (finger or numerical) representations should be drawn with caution. This would also be in line with the results of Lafay et al. (2013) who showed that with 4–7 year-olds finger gnosis was related to an enumeration task, but not to direct finger use in counting. In this context, Reeve and Humberstone (2011) have identified four subgroups of 5–7 year-old children based on their performance on an addition task and spontaneous finger use. In this classification, high performers rarely used their fingers directly, whereas moderate performers belonged to one of two groups: either to a group with high, or to a group with low, direct finger use. Finally, their fourth group contained low performance children and low finger use. In addition, Wasner et al. (2015) have shown for adults that the use of specific fingers can vary according to the demanded underlying principle of the task (e.g., requiring either ordinality or cardinality or 1-1 relations). This indicates that finger use is highly flexible and also depends on the task itself. Yet, training of finger gnosis and direct finger use in numerical tasks might have a double advantage for children. First, it may improve finger gnosis and finger representation itself. Second, it may help children to grasp the abstract format of numbers by using an embodied format of numbers (Moeller et al., 2012).

If numerical skills were rooted in finger representations, one would assume a universally applicable sequential development from using numerical gestures first to using abstract verbal numbers second. Piaget (1954) claimed that abstract concepts emerge from senso-motoric experiences. A study by Nicoladis et al. (2010) calls such a sequential development into question.

They showed that preschoolers were actually better at processing number words than at processing number gestures. Thus, at least for counting, they did not find number gestures to precede the use of symbolic number words. In a similar vein, Crollen et al. (2011a) have shown that blind and sighted 7–13 year-old children performed similarly in enumeration tasks despite less finger counting and more inconsistent finger-number associations on the part of the blind children (Crollen et al., 2011b). While both groups had equal finger discrimination abilities, blind children showed better working memory performance than sighted children. Thus, if finger counting facilitates the development of numerical skills in sighted children, then blind children might compensate for this effect with their superior working memory skills. This does not mean that finger counting cannot be useful (e.g., Lafay et al., 2013), especially for more complex and difficult tasks where finger counting could, for example, help to reduce working memory load (see also Crollen et al., 2011b). These studies suggest that although finger counting can be beneficial, it may not be necessary for developing counting abilities.

Intervention studies seem to be a promising tool to investigate whether there is a causal relation between finger gnosis, finger-number associations and arithmetic skills. Even though an increasing number of intervention studies have compared the contributions of potential precursor abilities for mathematic proficiency over the last years, only very few studies looked at the role of finger gnosis or finger-number association. To date, only a small number of studies have carried out finger-number trainings with school-aged children. For example, Gracia-Bafalluy and Noël (2008) provided a 30-min finger gnosis training session once a week, for an 8-week period, to first graders. The training was a 'pure' finger gnosis intervention designed to improve sensitivity and mobility of the fingers (e.g., labyrinth game or piano game). They observed that children with an initial poor finger gnosis benefited from the training and scored higher not only in finger gnosis, but also in numerical skills after the training. Unfortunately, their methodological procedure was rightfully criticized, because the authors did not consider the regression to the mean, which alternatively could explain the results (Fischer, 2010). In a recent study, Jay and Betenson (2017) trained 137 first graders in eight 30-min sessions during 4 weeks. The group playing finger gnosis games improved merely in the finger gnosis task. This is surprising, because in contrast to Gracia-Bafalluy and Noël (2008) their finger gnosis training involved not only 'pure' finger gnosis interventions, but also training in the cardinal and ordinal properties of numbers: Children actively verbalized numbers in games like finger counting, showing fingers-to-numbers or showing calculations with fingers. The group playing number games (e.g., domino, snake and ladders, playing with cards and dice) improved only in a non-symbolic magnitude comparison task. Finally, the third group, which had received a combination of both trainings, improved in their quantitative skills. The authors concluded that in the combined training children built up connections between different representations of numbers (e.g., finger-number, symbolic and non-symbolic representations), which might have led to the increased performance in quantitative skills compared to both single training groups.

Going beyond these two intervention studies Frey et al. (unpublished) trained 119 first graders not only in finger gnosis and finger counting, but also in using their fingers in arithmetic tasks in 18 sessions of approximately 25 min. Frey et al. (unpublished) trained the following skills: *Finger gnosis* was trained in the beginning of the intervention by differentiation and naming of the fingers, finger-thumb tapping and by tracing ways through labyrinths where children used each finger separately for finding different ways through various labyrinths. Further, children traced Arabic digits from 1 to 10 with their respective fingers or thumbs. *Ordinal number-finger association* was trained by a task asking children to count their fingers forward and backward thereby relating numbers to the respective finger. *Cardinal finger-number association* was trained, for example, by detecting numbers in a story. Here children had to indicate the numbers by showing their fingers. Further, they also played a memory card game with symbolic cards (digits), non-symbolic cards (points) and finger pattern cards featuring the numbers 1-9. Finally, most of the intervention games (nine tasks) trained *number relations* through the practice of addition and subtraction tasks while using the fingers (for a more detailed description of tasks see Frey, 2017). The results showed that trained children outperformed children of a control group in tasks including addition and subtraction up to a number range of 20, but not in number line estimation on a 0-to-50 and a 0-to-100 scale. Furthermore, these effects were still observed after 9 months. This study supports the view that training finger use in and beyond arithmetic tasks facilitates the learning of specific arithmetic skills. This does not necessarily mean that direct finger use while calculating increases the performance, but rather that the strengthening of the association between finger and number representations may lead to this improvement.

In sum, former studies have shown that primary school children improve in their arithmetic skills by finger-number training. However, some correlational studies suggest that finger-number relations might be predictors of later numerical skills and arithmetic already in preschoolers (Fischer et al., 2017; Suggate et al., 2017).

The aim of the present study is to investigate whether kindergarten children can profit from finger-number training, even before they receive formal math education in addition and subtraction. Training of other potential precursors has already been done (e.g., with non-symbolic approximate number training, Park et al., 2016; but see Szücs and Myers, 2017 for a critical review), but not with finger-number associations, to our knowledge. We are interested as to whether training finger-number associations in kindergarteners may pave the way for better future arithmetic skills as the training of phonological awareness paves the way for better future reading skills (e.g., Schneider et al., 1997; Bus and van Uzendoorn, 1999; Lundberg, 2009). To infer such a causal relation, it is important to train children before they receive formal instruction. For reading acquisition, this has been a debate for years: In school, literacy acquisition interacts with the acquisition of phonological awareness. Therefore, no clear conclusions about a causal relation can be drawn from children that already attend school (Castles and Coltheart, 2004). The same may also apply for the finger-number-arithmetic-relation examined here. The development of finger-number associations might interact with the acquisition of arithmetic proficiency.

To investigate whether finger-number associations can be trained in kindergarteners and whether this training affects arithmetic skills, we adapted the training of Frey et al. (unpublished) for kindergarten children aged five to six. An advantage in training younger children might be that for them finger representations might not be as mature as in older children. The same is true for finger-number associations, which may be less stable compared to older children. For that reason, both – finger representations and finger-number associations – might be even more susceptible to external training in younger compared to older children. That younger children might benefit more from interventions than older children has also been shown in other training studies with preschoolers (e.g., Park et al., 2016). In sum, we hypothesized that finger-number associations are causally related to numerical skills. If this is the case, then training of finger-number associations, especially in kindergarteners, may directly impact upon initial arithmetic performance – even before the beginning of formal arithmetic instruction and this impact should be larger than in a control group.

Although evidence for an influence on finger gnosis on later arithmetical performance seems rather small – if it exists at all – we incorporated some tasks of finger gnosis in the training, because finger gnosis seems to be necessary (but not sufficient) to associate fingers and numbers. In other words, if a child is not able to select or move a certain finger at all, they will also not be able to select this finger in associations with certain ordinal, cardinal or 1-1-finger-number relations. Thus, most tasks involving the assessment of active finger-number relation require some knowledge (i.e., here gnosis), of which fingers are to be involved in the task. Therefore, as a sub-hypothesis, we also wished to examine the question of whether finger gnosis at pre-test predicts initial mathematical skills at post-test.

However, it was not the aim of our study to show that training finger gnosis alone and unrelated to any finger-number relations has an effect on later arithmetic performance. We know that relations between finger gnosis and numerical skills are small to non-existent and have repeatedly argued (e.g., Domahs et al., 2010; Moeller et al., 2012) that the embodied representation of numbers with fingers, and not just finger gnosis alone, is essential.

Therefore, the core training feature concentrates on the finger-number associations as a precursor skill that might affect later arithmetic skills. However, we also include some early number relation tasks (completion to 5 and to 10) that may be on the border between finger-number associations and arithmetic skills (see section "Materials and Methods" for further details). Arithmetic knowledge of addition and subtraction were not directly trained, but they were accessed after the intervention. On purpose we decided to avoid training to the task because we wanted to investigate how the precursor skills of finger-number associations affect arithmetic skills without training arithmetical tasks by themselves.

In sum, the aim of this study was to investigate whether training finger-number associations in kindergarten improves initial arithmetic skills in elementary school. To the best of our knowledge, this is the first study that tries to show this causal relation by applying an intervention at kindergarten age with an active control group.

Here we wished to examine – as a first step – whether finger-number relations constitute a precursor of arithmetic skills, after taking into account an established predictor of early mathematical skills namely children's non-verbal intelligence (e.g., Aragón et al., 2016). In addition, we included gender as it is a debated popular predictor. In several studies gender differences have been observed in some spatial representations of number (e.g., Bull et al., 2013; Reinert et al., 2017), in children's early arithmetic skill (Krinzinger et al., 2012; Hornburg et al., 2017; see also Brunner et al., 2011), and even in adults' arithmetic and numerical skills (Pletzer et al., 2013, 2016). However, many recent studies have not found that females and males differ, for example, in a meta-analysis of math performance (Hyde, 2016), in several studies on children at various stages of their development (Morsanyi et al., 2018; Bakker et al., 2019; Hutchison et al., 2019); and in an adult online study testing the SNARC effect with over 1000 participants (see supplementary materials of Cipora et al., 2019). Because of these diverging results in the literature, which may differ depending on task, sample, culture and paradigm, we included gender as a predictor to examine whether it has any effect on embodied learning of basic numerical skills.

In sum, finger-number relations that were systematically targeting different constructs (finger gnosis, 1-1 finger-number mapping, ordinality, cardinality, base-10, place-value knowledge) were trained to increase salience of the training. If such a training in kindergarten were successful, future studies could investigate – in a second step – which components of finger-number relations might contribute the most to this training effect As a third step, further research can then compare or combine such a finger-number training with other effective interventions that train other components of numerical knowledge to unravel differential effects of the various potential trainings.

MATERIALS AND METHODS

Procedure

Preschool children received either finger-number training or one of two phonological control trainings. These phonological trainings belong to a training study on its own, but served as control training in the present study. The trainings were pseudorandomly assigned to local kindergartens to ensure that each training group comprised a similar number of children. For economical reasons, all children within the same kindergarten received the same training (but we tested whether there were pre-training differences between the kindergartens in the different experimental groups, which was not the case; see below). We allowed bilingual children to take part in the training, but only monolingual children were included in the study. Because our children were younger than the children in the study by Frey et al. (unpublished), we adapted the training's extent and content to suit kindergarteners. Each training session was only approximately 10 min, but the training took place every day, for a period of 10 weeks (from February/March to May/June during the children's final kindergarten year). Thus, the overall time of the training was nearly equal between our training study and that of Frey et al. (unpublished). In sum, we trained 18 groups of varying size (with a minimum of 4 children and a maximum of 10 children in the finger-number training). The training was conducted by instructed undergraduate students and doctorate members of the department of psychology of the University of Tübingen and took place in the kindergartens. Before and after the training we assessed each child's arithmetic and language skills in one or two test sessions lasting between 30 and 60 min. Tests that were important for the actual study included measures of finger gnosis, addition, subtraction and completion to 5/10. We also administered tests that were language specific to evaluate the phonological training. The results of the language study will be reported elsewhere.

Participants

In total 102 children took part in the training, and contributed data to both pre- and post-tests. The experimental group consisted of 35 children who received the finger-number training. The control group consisted of 67 children who received either the phonological training ($N = 37$, 23 male) or the phonological-orthographic training ($N = 30$, 17 male) as control trainings (see **Table 1** for demographic data)[1]. Participants received a present for each test session. Both children and their parents gave their informed consent. All children who took part in the tests were monolingual native speakers of German.

Materials and Tests

Training Material

The training material was adapted from Frey et al. (unpublished), and consisted of 18 different short games in total. We trained the following skills: *Finger gnosis* contained tasks like finger tapping and tracing a way through a labyrinth with specific fingers. Note, that these two tasks did not involve numbers. *1-to-1 mapping* of fingers and numbers included naming the fingers and mapping numbers to single fingers; learning Arabic digits was covered by tracing a number on a sheet with the respective fingers. *Ordinal finger-number associations* were trained by finger counting in various games (e.g., finger counting, object counting and counting of claps) and by ordering numbers, for example, by placing numbers in the right order and ordering a deck of cards displaying fingers, digits and points. The training of *cardinal finger-number associations* included games like naming the number corresponding to fingers presented, detecting numbers that were hidden in stories, playing a memory card game with cards displaying fingers and numbers, playing a bingo game with cards displaying fingers and sheets displaying numbers and playing a domino game with cards displaying fingers and numbers. Finally, *number relations* in *the base-10 and place-value system* and finger-number mapping were trained by completion of 5/10 tasks (one with fingers and one with a deck of cards displaying numbers) and by doubling numbers (showing double the number of fingers shown by the trainer). All games include the use of the fingers. In each training session up to three games were played depending on the length of the games (to see how often each game was played and for further details please refer to **Appendix Table A1**). The idea of having so many different games was not only to train different conceptual levels with increasing difficulty, but with 50 sessions it is also essential to vary the games to keep the children interested and motivated. The control training included phonological games of similar duration.

Pre- and Post-tests

Handedness

We used the lateralized quotient (LQ) of the Edinburgh inventory (Oldfield, 1971) to assess handedness, but we left out the item 'Striking Match.'

Finger gnosis

We used the same finger gnosis assessment as in Wasner et al. (2016) who adapted a task and procedure previously used by Noël (2005), Gracia-Bafalluy and Noël (2008), and Reeve and Humberstone (2011). For the first task, a box was placed over the hand of the child. The trainer touched a single finger on the middle phalanges and asked the child to show the tapped finger. This was done with both hands, respectively (maximum 6 points, 3 points for each hand). Thereafter, two fingers of one hand were touched consecutively. The child earned one point for each correct finger and another point for the correct order

[1] As suggested by a reviewer we ran all ANOVA and ANCOVA analyses including Bayes for each control group, separately (see **Supplementary Tables S8–S11** and **Supplementary Figure S1**). The results were similar to those when both control groups were merged into a single control group.

TABLE 1 | Demographic data and differences between groups in age, sex, attended days, handedness measured by the Lateralized Quotient (LQ; Oldfield, 1971) and in the subtest Matrices taken from the Culture Fair Intelligence Test (CFT 1-R; Weiß and Osterland, 2013).

	Age to pre-test [years; month (range)]	Sex (male/female)	Attended days [mean (SE, range)]	Handedness LG [mean (SE)]	Subtest Matrices [mean (SE)]
Experimental group	5.10 (53–6.11)	19/16	41.2 (1.15, 23.5–0)	70 (7.06)	5.8 (0.63)
Control group	5.11 (5.2–7)	40/27	40.0 (0.90, 7–49)	56 (6.54)	6.8 (0.42)
Significant differences between groups	$t < 1$, *ns*	$\chi 2 < 1$, *ns*	$t < 1$, *ns*	$t = 1.3$, $p = 0.154$, *ns*	$t = 1.3$, $p = 0.192$, *ns*

(maximum 20 points, 10 points for each hand). In the second task both hands were placed behind the box. Two pictures of the right and left hand were placed beside the box. The trainer touched one finger of the child and one finger of the picture at the same time. The child indicated whether the fingers were the same or not (4 points). Finally, children solved the same task, but with two fingers in succession (4 points). The maximum number of points was 34.

Completion-to-5/10

We introduced the completion-to-5 test with the following example: "Now I want you to tell me how many gummy bears we need to reach 5. If I have 4 gummy bears, how many more gummy bears do I need to reach 5?" A similar instruction served for the completion-to-10 task. The test stopped after 3 min. At pre-test, the maximum number of points was 15, and at post-test the maximum number of points was 30.

Addition

First, we familiarized the children with the concept of addition. At pretesting, children solved at maximum 25 tasks in the number range from 1 to 10. During post-testing, a maximum of 35 problems were presented (here the single numbers of the last 10 tasks ranged between 10 and 20). Children had 4 min to solve as many tasks as possible.

Subtraction

Again, we first familiarized the children with the concept of subtraction. At pretesting, children solved a maximum of 20 subtraction tasks in the number range of 1–10. At post-testing there were 30 problems. Thus, the maximum number of points was 30. Here, the numbers for the last five tasks ranged between 10 and 20. Again, the test stopped after 4 min.

General cognitive abilities

For a measure of general cognitive abilities, we administered two subtests (*Matrices* and *Continuing Rows*) of the Culture Fair Intelligence Test (CFT 1-R; Weiß and Osterland, 2013) at post-test. However, as various trainers reported that children had difficulties with the Continuing Rows subtest, we only entered the Matrices subtest into analyses.

All of the tasks were presented orally to the children and required a verbal response except the two tasks measuring general cognitive abilities where visual material was used in addition.

RESULTS

Each dependent measure (finger gnosis, completion, addition, subtraction) was subjected to a repeated measures ANOVA with the within-factor *Time* (pre-test versus post-test) and the two between-factors *Group* (experimental group versus control group) and *Sex* (male versus female) together with the co-variate *CFT-matrices*. The scores of the *CFT-matrices* were centered. **Figure 1** displays the mean scores of each dependent variable separately for each group and pre- and post-tests, respectively.

Independent *t*-tests showed that there was no hint of pre-test differences between experimental and control group for all tasks, $t_{all} \leq 0.731$, $p \geq 0.466$. All dependent measures showed that improvement took place over time implicating that the measures we used were sensitive to intra-individual changes.

Finger Gnosis

The ANOVA revealed a main effect of *Time*, $F(1,97) = 5.911$, $p = 0.017$, $\eta^2 = 0.056$. The co-variate *CFT-matrices* was also significant, $F(1,97) = 9.357$, $p = 0.003$, $\eta^2 = 0.088$. No other main effects or interactions were significant.

In order to quantify the null-effect of the interaction of interest (*Time* and *Group*) we applied Bayesian repeated measures ANOVA as implemented in JASP-software (JASP Team, 2017, Version 0.8.2). To get more assurance about the probability of the null hypothesis, we decided to run a Bayesian analysis. However, as there is no golden standard available, especially for repeated measures with within and between factors, we opted for the most simple and comprehensible way. We excluded Sex and CFT-matrices from the Bayesian analysis, because Sex was of no special interest here (and similarly distributed between groups) and CFT-matrices did not significantly differ between groups (see **Table 1**). We treated all main factors as nuisance factors to find out whether the interaction of interest (Time and Group) showed a higher probability for the null model or for the alternative model or whether it lay in between both models. The Bayes factor B01 indicates how much better the data predicts the null hypothesis compared to the alternative model. The detailed results of these analyses are provided in the **Supplementary Table S1**. For finger gnosis we set up a null model by excluding *CFT-matrices* and *Sex* and including each of the main factors (*Time* and *Group*) as nuisance variables. We compared this null model with an alternative model that included the interaction of interest (*Time* and *Group*). The model comparison revealed a BF_{01} of 5.06 for the interaction and a probability of $p(H_0| D) = 0.83$ which is substantial/positive evidence (Jarosz and Wiley, 2014) for the null model (see **Supplementary Material** for tables with Bayes Factors).

Each dependent post-measure (Finger gnosis, Completion, Addition and Subtraction) was additionally submitted to an ANCOVA with the fixed factors Group and Sex and the co-variates CFT-matrices and the respective pre-measure. Results of all ANCOVAs were nearly identical to the results of the ANOVAs (see **Supplementary Table S2** for detailed information).

Similar to the Bayesian repeated measures ANOVA we also ran Bayesian ANCOVAs. Here, we set up the null model by excluding CFT-matrices and Sex and by including the fixed factor Group. The co-variate Pre-measure was treated as a nuisance variable. We compared the null model with the alternative model that included the main effect of interest, namely Group. Results of Bayesian ANCOVAs were nearly identical to the Bayesian ANOVAs (see **Supplementary Table S3**).

Completion-to-5/10

The ANOVA revealed a main effect of *Time*, $F(1,97) = 47.616$, $p < 0.001$, $\eta^2 = 0.316$ and an effect of *CFT-matrices*, $F(1,97) = 23.105$, $p < 0.001$, $\eta^2 = 0.190$, and an interaction of

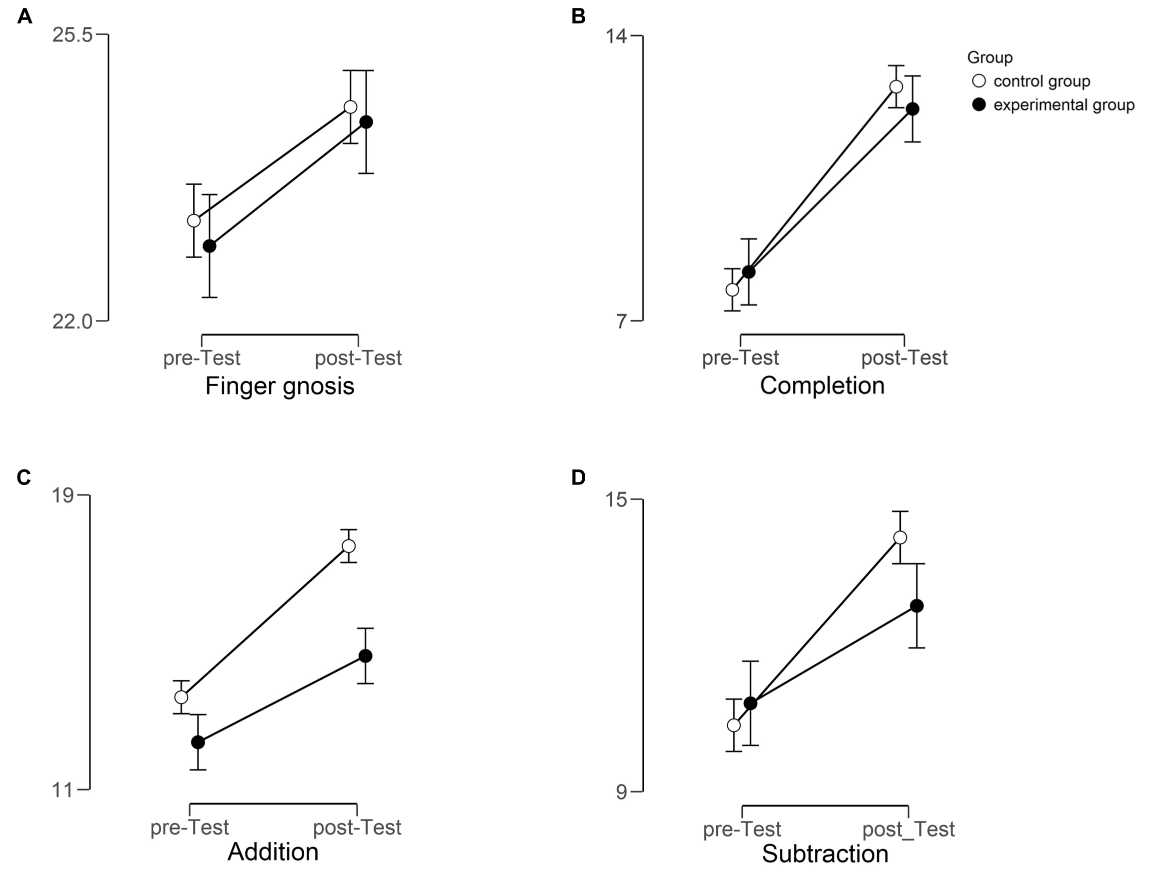


FIGURE 1 | Mean scores for all dependent variables [**(A)** Finger gnosis, **(B)** Completion, **(C)** Addition, and **(D)** Subtraction] for each group (experimental group [black] versus control group [white]) and each time (pre- versus post-test). Error bars indicate standard errors. Note, that the figure for subtraction displays results of the reduced sample (*N* = 63, for more details refer to the text).

both of these factors, $F(1,97) = 5.259$, $p = 0.024$, $\eta^2 = 0.035$. No other main effects or interactions were significant. Comparing the null model (excluding *CFT-matrices* and *Sex* and including the main factors *Time* and *Group* as nuisance variables) with the alternative model (including the interaction of *Time* and *Group)* revealed a BF_{01} of 3.56 for the interaction and a probability of 0.78 which is substantial/positive evidence for the null model.

Addition

The ANOVA revealed a main effect of *Time,* $F(1,97) = 29.748$, $p < 0.001$, $\eta^2 = 0.227$. Additionally, the covariate *CFT-matrices* was also significant, $F(1,97) = 28.983$, $p < 0.001$, $\eta^2 = 0.227$. No other main effects or interactions were significant. Comparing the null model (excluding *CFT-matrices* and *Sex* and including the main factors *Time* and *Group* as nuisance variables) with the alternative model (including the interaction of *Time* and *Group)* revealed a BF_{01} of 1.52 for the interaction and a probability of $p(H_0| D) = 0.60$ which is weak/anecdotal evidence for the null model. Thus, for addition there is no strong evidence either for the null model or for the alternative model.

Subtraction

Due to the fact that some children had profound difficulties in subtraction (some children were unable to solve even a single subtraction task), we excluded from analysis children who scored zero in pre- or post-tests. This reduced the original sample to 63 children ($N = 16$ in the experimental group, 5 female; $N = 47$ in the control group, 20 female). With this reduced sample, the ANOVA revealed a main effect of *Time,* $F(1,58) = 13.137$, $p < 0.001$, $\eta^2 = 0.181$. Additionally, we found an effect of *CFT-matrices*, $F(1,58) = 25.373$, $p < 0.001$, $\eta^2 = 0.290$. No other main effects or interactions were significant. Comparing the null model (excluding *CFT-matrices* and *Sex* and including the main factors *Time* and *Group* as nuisance variables) with the alternative model (including the interaction of *Time* and *Group)* revealed a BF_{01} of 1.64 for the interaction and a probability of $p(H_0| D) = 0.62$ which is weak/anecdotal evidence for the null model. Thus, similarly to addition, for subtraction there is no strong evidence either for the null model or for the alternative model.

Correlations

To characterize the relation between finger gnosis and arithmetic measures in more detail we calculated correlations and partial correlations (controlling for *CFT-matrices* scores) between all dependent measures pre- and post-test (see **Supplementary Tables S4, S5**). First, nearly all of our measures showed significant positive correlations pre- and post-test, respectively, as well as between pre- and post-test. This was supported by the Bayes-Factors indicating strong support for nearly all correlations compared to the null hypothesis (no correlation). However, correlations between arithmetic tasks (addition, subtraction and completion to 5/10) were consistently higher (0.61–0.79) than correlations between finger gnosis and arithmetic tasks (0.28–0.48) at pre- or post-test, respectively (see **Supplementary Table S4**).

Multiple Stepwise Regression

To examine whether finger gnosis at pre-test uniquely predicts any of the arithmetic skills at post-test beyond those at pre-test we ran a multiple stepwise regression. All predictors were taken from the pre-test. For addition at the post-test the final model included two predictors: addition and subtraction, $R^2 = 0.66$, $F(2,101) = 93.83$, $p < 0.001$. For subtraction the final model included three predictors: addition, subtraction and CFT, $R^2 = 0.54$, $F(3,101) = 38.39$, $p < 0.001$. For completion to 5/10 the final model included three predictors: addition, subtraction and completion to 5/10, $R^2 = 0.46$, $F(3,101) = 28.27$, $p < 0.001$. In sum, finger gnosis at pre-test did not significantly predict any dependent arithmetic measure at post-test, when other variables were included (see **Supplementary Table S6** for *Beta-* and *p*-values and **Supplementary Table S7** for Bayesian regression results). However, finger gnosis at pre-test did predict finger gnosis at post-test together with completion to 5/10, $R^2 = 0.25$, $F(2,101) = 16.74$, $p < 0.001$. Despite significance, the explained variance of the finger gnosis performance at post-test was lower than that of the other dependent measures at post-test.

DISCUSSION

This study sought to investigate whether combined finger-number training improves early arithmetic skills, even before formal arithmetic instruction has started. To this end, we provided training to 102 children in their final year of kindergarten. The training took place every day, for 10 min, for 10 weeks. An active control group of children received phonological training for identical duration and intensity. The results indicated that all children improved in their finger gnosis and arithmetic performance from pre- to post-test. However, this was independent of the training they received.

This outcome is surprising as Frey et al. (unpublished) showed robust effects of a similar finger-based training in first graders on tasks of addition and subtraction compared to an active control group. We discuss two possible groups of arguments for these findings; the first group referring to the possible inefficiency of the numerical intervention training, and the second referring to the possible efficiency of the non-numerical active control training. Specifically, first, we discuss arguments why the training may not have been successful for this particular age group with this particular training setting and for these particular evaluation tasks. Second, we discuss arguments why the control training contained elements (like implicitly training sequences) that might have been beneficial for elementary numerical and arithmetic tasks as well. Finally, we discuss the underlying reasoning of some of our intervention choices and how they affect the results and interpretation.

Reasons Why the Training Might Be Less Successful Than Other Finger-Number Trainings

Two of the dependent variables trained by Frey et al. (unpublished) were also directly trained in the present study: While finger gnosis training games differed from finger gnosis test items, the completion to 5/10 task was highly similar for training and tests. Nonetheless, Bayesian-Factor analysis revealed that the null model incorporating only the main effects seemed more probable compared to a model including the interaction of training group and time for both – finger gnosis and completion to 5/10 – measures. Unfortunately, Frey et al. (unpublished) tested neither finger gnosis nor completion to 5/10 at post-test, thus we cannot compare our outcomes in these measures with their training study in first graders. In contrast to our study, in the study by Gracia-Bafalluy and Noël (2008) only children in the 'pure' finger gnosis intervention group improved about 3.2 points in finger gnosis, but not children of the control intervention group. However, note that this effect could be due to a regression to the mean (Fischer, 2010) and might not be representative. Similarly, Jay and Betenson (2017) found a (small, but significant) increase of 1.9 points in finger gnosis only in groups receiving finger gnosis training. However, this rather small improvement might have been due to the combined group, because the authors analyzed both groups receiving finger gnosis training – single and combined group – together. It would be interesting to know whether the finger gnosis group and the combined group differed in their finger gnosis improvement. Note that in their study 'finger gnosis training' refers to activities that linked cardinal and ordinal properties of numbers to the fingers, i.e., they trained competencies like finger counting, finger-to-numbers relations or calculations with fingers. Thus, their training was comparable to ours. Yet, we found a similar improvement of 1.4 points for all groups, independently of the specific training. In addition, compared to the above-mentioned studies our children were on average 1 year younger – therefore, differences in training effects between those studies might also be attributable to the age and experience of the children.

Concerning the arithmetic measures, Jay and Betenson (2017) found that children receiving the combined training of finger gnosis and number games activities showed the largest gain in quantitative scores. While children from the other groups also improved in quantitative scores, their improvement was only half of that of the combined training group. Their quantitative score combined different measures. Some of these measures might be more related to the finger gnosis training (e.g., counting, adding dots on dice, splitting and combination of symbolic numbers); whereas others might be more related to the number training (e.g., ordering numbers, completion of number sequences, splitting and combination of non-symbolic numbers). It would have been interesting to see whether the finger gnosis group and the number group scored differently on subtasks combined in the quantitative score or whether children improved equally in all kinds of tasks from pre- to post-test. Indeed, the combined score might have obscured differential influences of finger gnosis (and number training) on different numerical skills. In contrast to the combined quantitative score of Gracia-Bafalluy and Noël (2008) and Jay and Betenson (2017) measured single numerical skills and children of the finger gnosis training improved in 'draw a hand' as well as in counting fingers, especially when larger number of fingers were involved (yet, improvement was only observed in response times, not in overall score). Finally, children improved in subitizing and ordinality score (comparing Arabic digits), but not in counting, magnitude comparison, enumeration and calculation. Thus, it might be that the influence of finger gnosis on numerical abilities comprises by far not all, but rather specific numerical skills.

Another important difference between the studies relates to the games that were trained. Moreover, these differences in training are related to the different levels of skills existing in the different age groups (kindergarteners versus primary school children). First, in the present study *addition and subtraction were not directly trained and combined with finger use* as in the study with the first graders (Frey et al., unpublished). The fact that direct training of tasks was successful in the study by Frey and colleagues is indirectly supported by the result of the number line accuracy task. Trained children showed no improvement in number line accuracy (Frey et al., unpublished). The authors argue this might be because the task is difficult to solve with the help of the fingers. Alternatively, this result could have emerged because number line accuracy was not practiced in the training; whereas addition and subtraction were directly trained. Now, turning to the *level of training*, most of the games trained in Frey's study on first graders covered number relations; whereas our training for kindergarten children included more games tapping into ordinality and cardinality. The different focuses of the trainings were also due to the fact that kindergarten children have a less stable quantity-number concept than first graders. Thus, the kindergarteners required and received more games involving the learning and understanding of the finger-number relations and Arabic numbers; whereas the first graders received more exercises in using their fingers directly in addition and subtraction tasks. Thus, kindergarteners received only a few tasks which directly trained actual arithmetic skills, such as the tasks completion to 5/10 or double numbers. Moreover, none of the tasks in our study explicitly trained addition or subtraction. In contrast, the first graders in the study by Frey et al. (unpublished) received instruction to use their fingers directly in various addition and subtraction games. Thus, we might have missed training the critical level or modules (e.g., finger use in arithmetic tasks) as intensively as in the case of the first graders in the study by Frey et al. (unpublished). However, as kindergarteners do not have the same numerical and arithmetic requirements as first graders, we deliberately concentrated more on *preceding* stages of finger-numerical development (e.g., finger counting, finger-number mapping). This concentration on early stages of finger-number development might have had less of an effect on actual arithmetic skills.

However, we made clear that the focus of our study was to see whether finger-number precursor training in kindergarten has positive effects on arithmetic skills (in a similar way, this has claimed for phonological awareness and later reading performance). The present study establishes that was not the case. We believe that this is important, because embodied training of numbers and in particular finger training has been advocated by ourselves and others (e.g., Moeller et al., 2012) as a means to improve early mathematic skills. This does not of course, either preclude that another form of finger-number training or other forms of precursor training (e.g., board games, or embodied spatial-numerical training, cf. Fischer et al., 2011), may have lasting training effects. A crucial question for the future is which training, which training setups or maybe which combinations of numerical/arithmetic intervention in kindergarten are most successful in training numerical/arithmetic precursor abilities in children.

Note, that we trained all children to use their fingers with corresponding numbers in the same way. Children were trained to start with the thumb of their right hand and count up to the pinkie. For the numbers 6–10 the same order of the fingers of the left hand was used. One issue raised by one reviewer, was that we might have "deconstructed" finger-number associations that may have been already constructed by children. Thus, our results may be negative due to the children in the experimental group who counted using a divergent finger pattern at pre-intervention. For Western adults, Lindemann et al. (2011) observed that 87.5% started to count with the thumb up to the pinkie and used the same sequence of fingers for the other hand. Thus, the finger counting sequence seems to be similar among most people. In contrast to the finger sequence, the starting hand seems to be more equally distributed (Lindemann et al., 2011). Moreover, studies have shown that the task used to collect the finger counting routines (e.g., questionnaire versus spontaneous use) influenced the outcome (e.g., Lucidi and Thevenot, 2014). For example. Wasner et al. (2014) showed that finger counting habits can change heavily according to situated circumstances. When the typical Fischer (2008) and Lindemann et al. (2011) finger counting questionnaire was administered about 54% reported counting from left-to-right. When participants additionally had a pencil in their hand, even more, 62% reported counting from left-to-right. When now the horizontally aligned finger picture used in the Lindemann questionnaire was removed and participants had to count spontaneously, the left-to-right advantage not only disappeared but even reversed. With empty hands and no picture of hands in front of them, the majority of people (72%) started from right to left. This shows that people are not fixed in their counting habits, but very flexible. Moreover, they also change their finger to number-relationships substantially depending on whether they refer to cardinal numbers, to ordinal numbers or to a 1-1 relationship between finger and number

(Wasner et al., 2015; which is the reason, why we trained all three of them). What is more, a recent study of Hohol et al. (2018) assessed the reliability and flexibility of finger counting habits. While reliability was satisfactory (about 75% reported using the same hand on both occasions), participants also reported huge flexibility. Overwhelmingly they said that they are also comfortable starting counting with the non-preferred hand, and about 50% even said that if they hold an object in their preferred starting hand, they do not bother to change hands or put the object away, but just start counting with the other hand. These studies point to a substantial flexibility in counting habits.

Nevertheless, because we tested kindergarten children, one might argue that they have less flexibility than first graders or adults tested in above studies (see Sato and Lalain, 2008; Previtali et al., 2011 for developmental data). Therefore, we reanalyzed all our data to see if there was any difference between children who were trained in congruence with their finger counting preference and those who were not. In the finger gnosis task children were asked to count to ten with their fingers. We compared two groups: one group who was trained in congruence with their preference, and the other group who was not. At pre-intervention, in the experimental group, 27 children counted in the *trained pattern* (in which 6 children switched to a divergent pattern at post-intervention), 8 children counted in a *divergent pattern* (in which 6 children switched to the trained counting pattern at post-intervention). The two groups did not differ in any of the post-tests (Mann–Whitney), $p_{(\text{finger gnosis, completion, addition})} > 0.65$. In the reduced sample for subtraction, 11 children with (pre-intervention) *trained counting pattern* and 5 children with (pre-intervention) *divergent counting pattern* were included. They did not differ in subtraction at post-test, $p = 0.69$. Obviously, the results have to be interpreted with caution, because of the different and small sample size, but, for the moment, there was no indication that the congruency of training direction with natural habits had an effect in any analysis. These data are consistent with the flexibility shown in the studies above and clearly inconsistent with the assumption that this issue affected training success.

Why the Null Effect Could Be Due to Improvement of the Control Intervention

One important difference between former finger training studies and our study is the *control intervention*. Frey et al. (unpublished) and Jay and Betenson (2017) had only no-intervention control groups. Gracia-Bafalluy and Noël (2008) had a story comprehension control group and a no-intervention control group. In contrast, we compared our finger-number training to a group trained in phonological awareness. Thus, domain-general factors might have improved with both kinds of trainings as well as domain-specific factors that might have overlapped in both training groups.

It is known that *domain-general variables* (e.g., concentration, attention, executive functions) can modulate performance in domain-specific skills (e.g., see Aunio and Niemivirta, 2010 how inattention modulated numerical performance). The influence of domain-general skills on specific skills might of course depend on the particular domain-general and/or domain-specific variable. For example, the causal relation between working memory and arithmetic skills is heatedly debated (Welsh et al., 2010; Melby-Lervag and Hulme, 2013; Cragg and Gilmore, 2014; Passolunghi and Costa, 2016; Honore and Noël, 2017; Ramani et al., 2017). Moreover, the strength of this relation may also depend on other factors, for example, whether children come from low-income families and/or whether children may have a risk for special impairments. Likewise, specific interventions (such as the training) provided to children in our study may have general effects on attention, concentration, motivation, working memory and other domains. Thus, what might have happened in our study is that the phonological training group was trained in general-domain variables and this, in turn, also led to improvement in their numerical skills (Purpura and Ganley, 2014; but see also Purpura and Reid, 2016).

Initially, we thought we had constructed our control trainings in such a way as there was no overlap in the training of specific skills (finger-number skills versus phonological awareness skills). However, taking a closer look at the specific exercises in both trainings may reveal certain similarities of trained *domain-specific factors*. Possible candidates are sequencing and ordinality, which both apply for numbers as well as for words (for example, one can count and order sounds in a spoken word). Thus, implicit training of these concepts in the phonological group might have generalized to the positive outcome in the numerical tasks. For example, one game in the phonological group involved counting a phoneme sequence in a word (e.g., M-U-MM-Y), which might have directly trained both ordinal-numerical as well as phoneme-skills.

This interpretation is supported by studies showing a relation between domain-general ordering skills (by using ordering of months or letters) and arithmetic skills in children (e.g., O'Connor et al., 2018) and adults (e.g., Morsanyi et al., 2017; Sasanguie et al., 2017; Vos et al., 2017).

A recent study of Xu and LeFevre (2016) show that learning sequential relations is beneficial for later arithmetic and numerical skills. It is therefore possible that more sequential finger-number games would have been beneficial for training success. Again, our non-numerical control training was also training sequential processes albeit not for numbers. As already discussed, children improved in both training conditions, the experimental and the active control training. Relating this to Xu and LeFevre (2016), one might suggest that in our control training, we have also trained sequential relations – although these relations were non-numerical, there might have been transfer effects to sequential numerical knowledge, which is an important cornerstone for later arithmetic skills. Note that in this respect our training lasted 10 weeks (Xu and LeFevre, 2016: 3 weeks), which leaves much time for implicit and explicit transfer effects.

Finger Gnosis Was Not a Predictor in This Study

Turning to the sub-question of whether finger gnosis is a predictor for later arithmetic skills, our regression result did not support this claim. Although finger gnosis correlated with arithmetic performance, it did not uniquely predict any of our arithmetic measures. These results are in line with studies that assume that factors other than finger gnosis – namely numerical knowledge and initial arithmetic abilities – might be more important in predicting later arithmetic skills (Long et al., 2016). Still, others have shown that finger gnosis can predict at least a small variance of later arithmetic performance (Penner-Wilger et al., 2007, 2009; Kohn et al., 2015; Poltz et al., 2015; Wyschkon et al., 2015; Wasner et al., 2016). However, a combination of the young age (leading to more error variance in the testing) and other control variables may be responsible for finger gnosis not being a predictor in the current study.

Active Control Group Rather Than Waiting List Control Group

We view as strength of our study that we used an active control group and not just a waiting control group. Note that in the child literature waiting control groups are viewed from critically to not acceptable (Fischer et al., 2013) and some authors do not include intervention studies without active control groups in their reviews (Slavin et al., 2009). The reason is that waiting control groups do not allow for the distinguishing of intervention-specific effects from intervention-unspecific effects such as attention, motivation or unspecific cognitive factors (learning how to learn) from intervention-specific effects, such as learning finger-number relations in our study. A recent meta-analysis confirmed this concern. Intervention studies without active control groups had generally larger effect sizes (Fischer et al., 2013). However, it is impossible to distinguish the contribution of intervention-unspecific and intervention-specific effects for such effect sizes. Therefore, we used an active control design and did not add a waiting control group, because it would not allow any substantial additional interpretation as regards the specific effects of our training.

Multiple Component vs. Single-Component Interventions

When one reviews intervention studies, it is essential to distinguish between short-term interventions, where one component in one game or task is trained, and long-term interventions, where multiple components and tasks are trained (see Fischer et al., 2013, for an overview). Some of us have conducted single-component embodied interventions targeting embodied numerosity in different variations (e.g., Fischer et al., 2011, 2015; Link et al., 2013; Link et al., 2014; Dackermann et al., 2016b; for reviews see Moeller et al., 2012, 2015; Dackermann et al., 2016a, 2017). When one conducts such trainings, it is inevitable that children get bored after a very short period of time. For instance, Fischer et al. (2015) could not even include post-tests after the second training in a cross-over design, because the decreased motivation of the children caused performance to drop substantially in the second post-test.

Any long-term intervention in such young children therefore necessarily cannot rely on one component, because it would get boring for the children after a few or even one session. We are not aware of any long-term intervention in numerical cognition which lasted over 50 sessions in 10 weeks (or more) and which used only one particular game for any numerical construct. All comparable interventions we are aware of used multiple modules and multiple games to improve one or more particular conceptual representation or process. Therefore, in any (not only our) long-term intervention with kindergarten children, it will always be impossible to track down any eventual changes to one particular game or module. This is only possible in short-term interventions with very few sessions, where children do not get bored by multiple repetitions of the same simple arithmetic game.

We have included finger gnosis in our multi-component finger-number intervention program, because earlier results (e.g., Noël, 2005; Wasner et al., 2016) suggested that finger gnosis may be weakly related to arithmetic skill. However, of our whole training modules, only two short training games exclusively targeted finger gnosis, all other games were explicitly related to finger-number relations. Thus, training finger gnosis was a very small part of the multi-component intervention program and given that the relations between finger gnosis and math are weak, we do not believe that their inclusion had a large impact on the results. However, theoretically, we cannot preclude that these two of the 18 games contributed to the null effects in this study.

Limitations

As the finger gnosis and finger-number training provided in the current study obviously was not effective beyond the control group, it might be that the training ought to be *provided together with formal arithmetic instruction*. A key difference between the kindergarten children in our study and the first graders in the study by Frey et al. (unpublished) was that the latter had already been formally introduced to the concept of addition and subtraction at school, which of course was not the case for our kindergarten children. The lack of formal arithmetic education did not prevent some of the kindergarten children from solving quite a few of the addition tasks. On the other hand, the subtraction tasks were very difficult and often frustrating for nearly all of the children. The latter was also obvious as this task showed a high fluctuation in performance. Nearly 20 percent of the children could not solve even one of the subtraction items at post-test, but the same children had solved an average of nearly five items at pre-test. This observation might be a consequence of the fact that kindergarteners are used to counting forward rather than backward. In line with this, it has been shown that preschoolers had more difficulties using a task to access the preceding compared to the next number (e.g., Sella et al., 2019; Sella and Lucangeli, 2020). In addition, the fact that basic arithmetic performance varies strongly at this age may be due to fluctuations in attention and motivation (see, e.g., Aunio and Niemivirta, 2010 for the influence of inattention on arithmetic performance). Other studies have also found large individual differences in numerical abilities in preschoolers (e.g., Weinhold Zulauf et al., 2003; Dowker, 2008).

Further, as we did not control for *external interventions* taking place at the individual kindergarten or at the homes of the children, it might be that these interventions leveled out the effects of the training. However, although often not mentioned this applies to all of the studies in the field, since no kindergarten, school and probably almost no parents would agree to participate in a longitudinal numerical study in which all numerical/arithmetic activities are forbidden for the time of the longitudinal study this might create an additional source of error variance. Additionally, stronger promotion of numerical skills in the kindergarten and/or in the family might in turn also boost numerical knowledge. This may be even reinforced by the fact that in and around the city of Tübingen, where the training took place, families have an above-average socioeconomic status, and thus, children may have been promoted even more. If many of the children in our study received a great deal of such numerical promotion in their kindergartens or families anyway in this developmental period, this could have prevented our training from having a visible additional benefit. Thus, the training may still be beneficial for (possibly lower SES) families, in which numerical skills of children are supported or promoted to a lesser extent.

It could also be the case when familiarizing kindergarteners with numbers the *increased interest* in one domain might generalize for neighboring domains like sounds and letters and vice versa, thereby promoting improvement in both fields. The finding that numerical skills obviously improve dramatically during the last kindergarten year was also shown by Weinhold Zulauf et al. (2003) who tested over 300 German-speaking children in Austria (see also Krajewski et al., 2008). The authors even speak of a *"sensitive period"* for the acquisition of numerical skills. Thus, children at this age gain knowledge in the domain of numbers very fast through natural interest.

At last, we do not want to omit the possibility that the training might have had no effect whatsoever. In this case, *overall maturation*, which is certainly fast at that age, might have led to the improvement of all skills in all groups. However, we do not think maturation plays a sole role, as other studies with waiting control groups consistently showed differences when compared to the intervention groups (e.g., Gracia-Bafalluy and Noël, 2008; Jay and Betenson, 2017; Frey et al., unpublished). Moreover, other studies focusing on other numerical precursor skills, or including a broader range of such skills, have shown intervention effects in kindergarten children (e.g., Kaufmann et al., 2005; Krajewski et al., 2008; Praet and Desoete, 2014).

CONCLUSION AND PERSPECTIVES

In sum, we suggest that the difference in training and age was responsible for the different outcomes between the Frey et al. (unpublished) study and the current study. The first graders in the study by Frey et al. (unpublished) had received training in number relations and direct finger use for addition and subtraction, and the kindergarten children in our study had received training in a quantity of number concepts. Both studies trained a variety of different skills occurring at different developmental stages (finger gnosis, 1-1 finger-number mapping, ordinality and cardinality of numbers and number relations in base-10 and place-value system). It may be a rather complicated but potentially rewarding task for future studies to try and disentangle these factors and test more directly which specific components of the training were responsible for the training effect in Frey et al. (unpublished) first-graders and which components might be more promising for training in kindergarteners compared to older children.

Maybe one should also take the developmental stage of the individual child into account. For example, it might be fruitful to apply an adaptive finger-based numerical training suited to the needs of the individual child (similar to, e.g., Praet and Desoete, 2014, for computerized counting), rather than having all children play the same games. Given the large individual differences in preschooler's numerical abilities (Dowker, 2008), a lot of the games might be boring for some children but overburden others. Individual interventions carried out in primary school directly trained weak number skills of individual children (e.g., Dowker and Sigley, 2010; Holmes and Dowker, 2013). The individual arithmetical skills of the children trained in these studies were highly susceptible to the individual intervention. Some of the concepts used in the training were similar to ours (e.g., counting, written symbols, etc.), whereas others tapped more into conceptual and reasoning domains. Thus, again by comparing these interventions in primary school with our kindergarten training it is difficult to uncover the effective (or ineffective) components of our training. Differences of outcome could also be due to the different characteristics of the groups (preschool-aged normally developing children versus school-aged children with arithmetic difficulties). In sum, different outcomes could be due to the different trainings, the trained skills, or the individual adaption of the training. Finally, it could be due to a combination of all three factors. Thus, it remains for future research to find out whether, and what components of, finger-based numerical training can be trained at which ages (specifically kindergarten versus primary school) and which training might be best-suited for normally developing or at-risk children (see Kaufmann et al., 2003; Dowker and Sigley, 2010; Holmes and Dowker, 2013 for interventions in primary school children with arithmetic difficulties). Moreover, a comparison and/or combination of finger-based numerical training with other components, that have been found to be effective, e.g. conceptual training (for kindergarten children see Kaufmann et al., 2005) might be fruitful.

CONCLUSION

All of our kindergarteners showed improved scores in finger-gnosis, addition, subtraction and completion to 5/10, independent of the training they received. We argue that these general improvements could have been due to both domain-general and domain-specific training effects. As our control training contained elements (such as sequencing or ordinality) that might have been beneficial for numerical skills as well a final evaluation of the training as being effective or ineffective is preliminary and may require a different active control group. Further studies investigating how finger-number trainings in kindergarten children might affect the development of numerical skills should incorporate different active intervention control groups to disentangle general and specific training effects from maturation effects and environmental factors like institutional or private promotion. Finally, as a first intervention study where finger-number associations were trained in normally developing kindergarteners, our data provide insights about the impact of finger-number associations for arithmetic development. Even though we are convinced that appropriate embodied trainings might help (e.g., Dackermann et al., 2017), it is in our view important to also publish and acknowledge the limitations of such training approaches when they were not as successful as we would have ourselves postulated before we saw the data.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethical Committee of the German Psychological Association ("Ethikkommission der Deutschen Gesellschaft für Psychologie", US 082014). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

US conceived the study and took primary responsibility for drafting the manuscript. US, AB, and H-CN contributed to design of the training and the tasks. AB conducted the study. US analyzed the data. AB and H-CN commented on drafts. All authors contributed to the manuscript revision, read and approved the submitted version.

FUNDING

The work was supported by a grant of the German Research Foundation (SCHI 1335/1-1). We acknowledge support by Open Access Publishing Fund of University of Tübingen.

ACKNOWLEDGMENTS

We are grateful to Miriam Frey for helping us in adapting her finger-based training (Frey et al., unpublished) for kindergarten children. We are also grateful to Janina Althaus, Nina Bee, Sabine Braasch, Jana Denkinger, Lena Flik, Eva Franz, Stella Gellner, Isolde Kolb, Ruth Kessler, Janina Krempel, Johanna Lechner, Larissa Lorch, Monika Peterová Jarošová, Anne Rau, Angelika Schmitt, Anna-Lena Schultz, Daniel-Niels Tönsing, Gülsan Usta, and Marlene Weller for assistance in the training and in collecting the data. We warmly thank all kindergartens, parents, and children that participated in our study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00529/full#supplementary-material

REFERENCES

Aragón, E., Navarro, J. I., Aguilar, M., Cerda, G., and García-Sedeño, M. (2016). Predictive model for early math skills based on structural equations. *Scand. J. Psychol.* 57, 489–494. doi: 10.1111/sjop.12317

Aunio, P., and Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learn. Individ. Differ.* 20, 427–435. doi: 10.1016/j.lindif.2010.06.003

Bakker, M., Torbeyns, J., Wijns, N., Verschaffel, L., and De Smedt, B. (2019). Gender equality in 4- to 5-year-old preschoolers' early numerical competencies. *Devel. Sci.* 22:e12718. doi: 10.1111/desc.12718

Berteletti, I., and Booth, J. R. (2015). Perceiving fingers in single-digit arithmetic problems. *Front. Psychol.* 6:226. doi: 10.3389/fpsyg.2015.00226

Brozzoli, C., Ishihara, M., Göbel, S. M., Salemme, R., Rossetti, Y., and Farnè, A. (2008). Touch perception reveals the dominance of spatial over digital representation of numbers. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5644–5648. doi: 10.1073/pnas.0708414105

Brunner, M., Krauss, S., and Martignon, L. (2011). Eine alternative Modellierung von Geschlechtsunterschieden in Mathematik. *J. für Mathematik-Didaktik* 32, 179–204. doi: 10.1007/s13138-011-0026-2

Bull, R., Cleland, A. A., and Mitchell, T. (2013). Sex differences in the spatial representation of number. *J. Exp. Psychol.* 142, 181–192. doi: 10.1037/a0028387

Bus, A., and van Uzendoorn, M. H. (1999). Phonological awareness and early reading: a meta-analysis of experimental training studies. *J. Educ. Psychol.* 91, 403–414. doi: 10.1037/0022-0663.91.3.403

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

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

Cipora, K., Patro, K., and Nuerk, H.-C. (2015). Are spatial-numerical associations a cornerstone for arithmetic learning? The lack of genuine correlations suggests no. *Mind Brain and Educ.* 9, 190–206. doi: 10.1111/mbe.12093

Cipora, K., Soltanlou, M., Reips, U.-D., and Nuerk, H.-C. (2019). The SNARC and MARC effects measured online: large-scale assessment methods in flexible cognitive effects. *Behav. Res. Methods* 51, 1676–1692. doi: 10.3758/s13428-019-01213-1215

Costa, A. J., 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

Cragg, L., and Gilmore, C. (2014). Skills underlying mathematics: The role of executive function in the development of mathematics proficiency. *Trends Neurosci. Educ.* 3, 63–68. doi: 10.1016/j.tine.2013.12.001

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. doi: 10.1016/j.jecp.2011.03.011

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

Crollen, V., and Noël, M.-P. (2015). The role of fingers in the development of counting and arithmetic skills. *Acta Psychol.* 156, 37–44. doi: 10.1016/j.actpsy.2015.01.007

Crollen, V., NoëL, M.-P., Seron, X., Mahau, P., Lepore, F., and Collignon, O. (2014). Visual experience influences the interactions between fingers and numbers. *Cognition* 133, 91–96. doi: 10.1016/j.cognition.2014.06.002

Dackermann, T., Fischer, U., Cress, U., Nuerk, H. C., and Moeller, K. (2016a). Bewegtes lernen numerischer kompetenzen [embodied learning of numerical competencies]. *Psychol. Rundsch.* 67, 102–109. doi: 10.1026/0033-3042/a000302

Dackermann, T., Fischer, U., Huber, S., Nuerk, H. C., and Moeller, K. (2016b). Training the equidistant principle of number line spacing. *Cogn. Process.* 17, 243–258. doi: 10.1007/s10339-016-0763-8

Dackermann, T., Fischer, U., Nuerk, H.-C., Cress, U., and Moeller, K. (2017). Applying embodied cognition - From useful interventions and their theoretical underpinnings to practical applications. *ZDM Math. Educ.* 49, 545–557. doi: 10.1007/s11858-017-0850-z

De Smedt, B., Noël, M.-P., Gilmore, C., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills 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

Di Luca, S., and Pesenti, M. (2011). Finger numeral representations: more than just another symbolic code. *Front. Psychol.* 2:272. doi: 10.3389/fpsyg.2011.00272

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

Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Devel. Sci.* 11, 650–654. doi: 10.1111/j.1467-7687.2008.00713.x

Dowker, A., and Sigley, G. (2010). Targeted interventions for children with arithmetical difficulties. *Br. J. Educ. Psychol. Monogr.* 11, 65–81. doi: 10.1348/97818543370009X12583699332492

Fayol, M., Barrouillet, P., and Marinthe, C. (1998). Predicting arithmetical achievement from neuro-psychological performance: a longitudinal study. *Cognition* 68, B63–B70. doi: 10.1016/S0010-0277(98)00046-8

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. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex* 44, 386–392. doi: 10.1016/j.cortex.2007.08.004

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

Fischer, U., Moeller, K., Cress, U., and Nuerk, H.-C. (2013). Interventions Supporting Children's Mathematics School Success A Meta-Analytic Review. *Eur Psychol.* 18, 89–113. doi: 10.1027/1016-9040/a000141

Fischer, U., Moeller, K., Huber, S., Cress, U., and Nuerk, H.-C. (2015). Full-body movement in numerical trainings: a pilot study with an interactive whiteboard. *Int. J. Ser. Games* 2, S23–S35. doi: 10.17083/ijsg.v2i4.93

Fischer, U., Suggate, S. P., Schmirl, J., and Stoeger, H. (2017). Counting on fine motor skills: links between preschool finger dexterity and numerical skills. *Devel. Sci.* 21:e12623. doi: 10.1111/desc.12623

Frey, M. (2017). *Die Bedeutsamkeit Fingerbasierter Repräsentationen auf Numerische Fähigkeiten – Ergebnisse einer Intervention zu Beginn der Grundschule und Differenzierung Struktureller Merkmale*. Doctoral dissertation, Eberhard Karls Universität Tübingen, Germany.

Gerstmann, J. (1940). Syndrome of finger agnosia, disorientation for right and left, agraphia and acalculia: local diagnostic value. *Arch. Neurol. Psychiatr.* 44, 398–408. doi: 10.1001/archneurpsyc.1940.02280080158009

Gracia-Bafalluy, M., and Noël, M.-P. (2008). Does finger training increase young children's numerical performance? *Cortex* 44, 368–375. doi: 10.1016/j.cortex.2007.08.020

Hohol, M., Wołoszyn, K., Nuerk, H., and Cipora, K. (2018). A large-scale survey on finger counting routines, their temporal stability and flexibility in educated adults. *PeerJ* 6:e5878. doi: 10.7717/peerj.5878

Holmes, W., and Dowker, A. (2013). Catch Up Numeracy: a targeted intervention for children who are low-attaining in mathematics. *Res. Math. Educ.* 15, 249–265. doi: 10.1080/14794802.2013.803779

Honore, N., and Noël, M.-P. (2017). Can working memory training improve preschoolers'; numerical abilities? *J. Numer. Cogn.* 3, 516–539. doi: 10.5964/jnc.v3i2.54

Hornburg, C. B., Rieber, M. L., and McNeil, N. M. (2017). An integrative data analysis of gender differences in children's understanding of mathematical equivalence. *J. Exp. Child Psychol.* 163, 140–150. doi: 10.1016/j.jecp.2017.06.002

Hutchison, J. E., Lyons, I. M., and Ansari, D. (2019). More similar than different: gender differences in children's basic numerical skills are the exception not the rule. *Child Devel.* 90, e66–e79. doi: 10.1111/cdev.13044

Hyde, J. S. (2016). Sex and cognition: gender and cognitive functions. *Curr. Opin. Neurobiol.* 38, 53–56. doi: 10.1016/j.conb.2016.02.007

Jarosz, A. F., and Wiley, J. (2014). What are the odds? A practical guide to computing and reporting bayes factors. *J. Probl. Solving* 7, 2–9. doi: 10.7771/1932-6246.1167

JASP Team (2017). *JASP Version 0.8.1.2. Computer software*.

Jay, T., and Betenson, J. (2017). Mathematics at your fingertips: testing a finger training intervention to improve quantitative skills. *Front. Educ.* 2, 1–8. doi: 10.3389/feduc.2017.00022

Kaufmann, L., Delazer, M., Pohl, R., Semenza, C., and Dowker, A. (2005). Effects of a specific numeracy educational program in kindergarten children: a pilot study. *Educ. Res. Eval.* 11, 405–431. doi: 10.1080/13803610500110497

Kaufmann, L., Handl, P., and Thony, B. (2003). Evaluation of a numeracy intervention program focusing on basic numerical knowledge and conceptual knowledge: a pilot study. *J. Learn. Disabil.* 36, 564–573. doi: 10.1177/00222194030360060701

Kohn, J., Kucian, K., Wuithschick, E., Mayer, V., Rauscher, L., McCaskey, U., et al. (2015). Rechenleistung und fingergnosie: besteht ein zusammenhang? *Lernen und Lernstörungen* 4, 209–223. doi: 10.1024/2235-0977/a000106

Krajewski, K., Nieding, G., and Schneider, W. (2008). Kurz- und langfristige effekte mathematischer frühförderung im kindergarten durch das programm "mengen, zählen, zahlen". *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie* 40, 135–146. doi: 10.1026/0049-8637.40.3.135

Krinzinger, H., Kaufmann, L., Gregoire, J., Desoete, A., Nuerk, H. C., and Willmes, K. (2012). Gender differences in the development of numerical skills in four european countries. *Int. J. Gen. Sci. Technol.* 4, 62–77.

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

Libertus, M. E., Odic, D., Feigenson, L., and Halberda, J. (2016). The precision of mapping between number words and the approximate number system predicts children's formal math abilities. *J. Exp. Child Psychol.* 150, 207–226. doi: 10.1016/j.jecp.2016.06.003

Lindemann, O., Alipour, A., and Fischer, M. H. (2011). Finger counting habits in middle eastern and western individuals: an online survey. *J. Cross Cult. Psychol.* 42, 566–578. doi: 10.1177/0022022111406254

Link, T., Moeller, K., Huber, S., Fischer, U., and Nuerk, H.-C. (2013). Walk the number line - an embodied training of numerical concepts. *Trends Neurosci. Educ.* 2, 74–84. doi: 10.1016/j.tine.2013.06.005

Link, T., Schwarz, E. J., Huber, S., Fischer, U., Nuerk, H.-C., Cress, U., et al. (2014). Mathe mit der Matte - Verkörperlichtes Training basisnumerischer Kompetenzen. *Zeitschrift für Erziehungswissenschaft* 17, 257–277. doi: 10.1007/s11618-014-0533-2

Long, I., Malone, S. A., Tolan, A., Burgoyne, K., Heron-Delaney, M., Witteveen, K., et al. (2016). The cognitive foundations of early arithmetic skills: it is counting and number judgment, but not finger gnosis, that count. *J. Exp. Child Psychol.* 152, 327–334. doi: 10.1016/j.jecp.2016.08.005

Lucidi, A., and Thevenot, C. (2014). Do not count on me to imagine how I act: behavior contradicts questionnaire responses in the assessment of finger counting habits. *Behav. Res. Methods* 46, 1079–1087. doi: 10.3758/s13428-014-0447-1

Lundberg, I. (2009). Early precursors and enabling skills of reading acquisition. *Scand. J. Psychol*. 50, 611–616. doi: 10.1111/j.1467-9450.2009.00778.x

Melby-Lervag, M., and Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Dev. Psychol.* 49, 270–291. doi: 10.1037/a0028228

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), S271–S274. doi: 10.1007/s10339-012-0457-9

Moeller, K., Fischer, U., Nuerk, H.-C., and Cress, U. (2015). Computers in mathematics education - Training the mental number line. *Comput. Hum. Behav.* 48, 597–607. doi: 10.1016/j.chb.2015.01.048

Moeller, K., Martignon, L., Wessolowski, S., Engel, J., and Nuerk, H.-C. (2011). Effects of finger counting on numerical development – The opposing views of neurocognition and mathematics education. *Front. Psychol.* 2:328. doi: 10.3389/fpsyg.2011.00328

Moeller, K., and Nuerk, H.-C. (2012). Zählen und rechnen mit den fingern. *Lernen und Lernstörungen* 1, 33–53. doi: 10.1024/2235-0977/a000004

Morsanyi, K., Mahony, E. O., and Mccormack, T. (2017). Number comparison and number ordering as predictors of arithmetic performance in adults?: exploring the link between the two skills, and investigating the question of domain-specificity. *Q. J. Exp. Psychol.* 70, 2497–2517. doi: 10.1080/17470218.2016.1246577

Morsanyi, K., van Bers, B. M. C. W., McCormack, T., and McGourty, J. (2018). The prevalence of specific learning disorder in mathematics and comorbidity with other developmental disorders in primary school-age children. *Br J. Psychol.* 109, 917–940. doi: 10.1111/bjop.12322

Nguyen, T., Watts, T. W., Duncan, G. J., Clements, D. H., Sarama, J. S., Wolfe, C., et al. (2016). Which preschool mathematics competencies are most predictive of fifth grade achievement? *Early Child. Res. Q.* 36, 550–560. doi: 10.1016/j.ecresq.2016.02.003

Nicoladis, E., Pika, S., and Marentette, P. (2010). Are number gestures easier than number words for preschoolers? *Cogn. Dev.* 25, 247–261. doi: 10.1016/j.cogdev.2010.04.001

Noël, M. P. (2005). Finger gnosia: a predictor of numerical abilities in children? *Child Neuropsychol.* 11, 413–430. doi: 10.1080/09297040590951550

O'Connor, P. A., Morsanyi, K., and McCormack, T. (2018). Young children's non-numerical ordering ability at the start of formal education longitudinally predicts their symbolic number skills and academic achievement in maths. *Dev. Sci.* 21:e12645. doi: 10.1111/desc.12645

Oldfield, R. C. (1971). The assessment and analysis of handedness: the edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Park, J., Bermudez, V., Roberts, R. C., and Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *J. Exp. Child Psychol.* 152, 278–293. doi: 10.1016/j.jecp.2016.07.011

Passolunghi, M. C., and Costa, H. M. (2016). Working memory and early numeracy training in preschool children. *Child Neuropsychol.* 22, 81–98. doi: 10.1080/09297049.2014.971726

Penner-Wilger, M., Fast, L., LeFevre, J., Smith-Chant, B. L., Skwarchuk, S., Kamawar, D., et al. (2007). "The foundations of numeracy: Subitizing, finger gnosia, and fine-motor ability," in *Proceedings of the 29th Annual Cognitive Science Society*, (Ottawa: Carleton University), 1385–1390.

Penner-Wilger, M., Fast, L., LeFevre, J., Smith-Chant, B. L., Skwarchuk, S., Kamawar, D., et al. (2009). "Subitizing, finger gnosis, and the representation of number," in *Proceedings of the 31st Annual Cognitive Science Society*, (Ottawa: Carleton University), 520–525.

Piaget, J. (1954). *The Construction of Reality in the Child*. New York: Basic Books.

Pletzer, B., Kronbichler, M., Nuerk, H.-C., and Kerschbaum, H. (2013). Sex differences in the processing of global vs. local stimulus aspects in a two-digit number comparison task - an fMRI study. *PLoS One* 8:e53824. doi: 10.1371/journal.pone.0053824

Pletzer, B., Moeller, K., Scheuringer, A., Domahs, F., Kerschbaum, H. H., and Nuerk, H.-C. (2016). Behavioral evidence for sex differences in the overlap between subtraction and multiplication. *Cogn. Process.* 17, 147–154. doi: 10.1007/s10339-016-0753-x

Poeck, K. (1964). Phantoms following amputation in early childhood and in congenital absence of limbs. *Cortex* 1, 269–275. doi: 10.1016/S0010-9452(64)80002-2

Poltz, N., Wyschkon, A., Höse, A., Aster, M. V., and Esser, G. (2015). Vom Fingergefühl zum Rechnen. *Lernen und Lernstörungen* 4, 177–193. doi: 10.1024/2235-0977/a000088

Praet, M., and Desoete, A. (2014). Enhancing young children's arithmetic skills through non-intensive, computerised kindergarten interventions: a randomised controlled study. *Teach. Teach. Educ.* 39, 56–65. doi: 10.1016/j.tate.2013.12.003

Previtali, P., Rinaldi, L., and Girelli, L. (2011). Nature or nurture in finger counting: a review on the determinants of the direction of number-finger mapping. *Front. Psychol.* 2:363. doi: 10.3389/fpsyg.2011.00363

Purpura, D. J., Daya, E., Napolia, A. R., and Hart, S. A. (2017). Identifying domain-general and domain-specific predictors of low mathematics performance: a alassification and regression tree analysis. *J. Numer. Cogn.* 3, 365–399. doi: 10.5964/jnc.v3i2.53

Purpura, D. J., and Ganley, C. M. (2014). Working memory and language: skill-specific or domain-general relations to mathematics? *J. Exp. Child Psychol.* 122, 104–121. doi: 10.1016/j.jecp.2013.12.009

Purpura, D. J., and Reid, E. E. (2016). Mathematics and language: individual and group differences in mathematical language skills in young children. *Early Child. Res. Q.* 36, 259–268. doi: 10.1016/j.ecresq.2015.12.020

Ramani, G. B., Jaeggi, S. M., Daubert, E. N., and Buschkuehl, M. (2017). Domain-specific and domain-general training to improve kindergarten children's mathematics. *J. Numer. Cogn.* 3, 468–495. doi: 10.5964/jnc.v3i2.31

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

Reinert, R. M., Huber, S., Nuerk, H.-C., and Moeller, K. (2017). Sex differences in number line estimation: the role of numerical estimation. *Br. J. Psychol.* 108, 334–350. doi: 10.1111/bjop.12203

Sasanguie, D., Lyons, I. M., De Smedt, B., and Reynvoet, B. (2017). Unpacking symbolic number comparison and its relation with arithmetic in adults. *Cognition* 165, 26–38. doi: 10.1016/j.cognition.2017.04.007

Sato, M., and Lalain, M. (2008). On the relationship between handedness and hand-digit mapping in finger counting. *Cortex* 44, 393–399. doi: 10.1016/j.cortex.2007.08.005

Schneider, W., Küspert, P., Roth, E., Vise, M., and Marx, H. (1997). Short- and long-term effects of training phonological awareness in kindergarten: evidence from two German studies. *J. Exp. Child Psychol.* 66, 311–340. doi: 10.1006/jecp.1997.2384

Sella, F., and Lucangeli, D. (2020). The knowledge of the preceding number reveals a mature understanding of the number sequence. *Cognition* 194:104104. doi: 10.1016/j.cognition.2019.104104

Sella, F., Lucangeli, D., Cohen Kadosh, R., and Zorzi, M. (2019). Making sense of number words and arabic digits: does order count more? *Child Dev.* doi: 10.1111/cdev.13335 [Epub ahead of print],

Slavin, R. E., Lake, C., and Groff, C. (2009). Effective programsin middle and high school mathematics: a best-evidencesynthesis. *Rev. Educ. Res.* 79, 839–911. doi: 10.3102/0034654308330968

Suggate, S., Stoeger, H., and Fischer, U. (2017). Finger-based numerical skills link fine motor skills to numerical development in preschoolers. *Percept. Mot. Skills* 124, 1085–1106. doi: 10.1177/0031512517727405

Szücs, D., and Myers, T. (2017). A critical analysis of design, facts, bias and inference in the approximate number system training literature: a systematic review. *Trends Neurosci. Educ.* 6, 187–203. doi: 10.1016/j.tine.2016.11.002

Tschentscher, N., Hauk, O., Fischer, M. H., and Pulvermüller, F. (2012). You can count on the motor cortex: finger counting habits modulate motor cortex activation evoked by numbers. *NeuroImage* 59, 3139–3148. doi: 10.1016/j.neuroimage.2011.11.037

Vos, H., Sasanguie, D., Gevers, W., and Reynvoet, B. (2017). The role of general and number-specific order processing in adults' arithmetic performance. *J. Cogn. Psychol.* 29, 469–482. doi: 10.1080/20445911.2017.1282490

Wasner, M., Moeller, K., Fischer, M. H., and Nuerk, H.-C. (2014). Aspects of situated cognition in embodied numerosity: the case of finger counting. *Cogn. Process.* 15, 317–328. doi: 10.1007/s10339-014-0599-z

Wasner, M., Moeller, K., Fischer, M. H., and Nuerk, H.-C. (2015). Related but not the same: ordinality, cardinality and 1-to-1 correspondence in finger-based numerical representations. *J. Cogn. Psychol.* 27, 426–441. doi: 10.1080/20445911.2014.964719

Wasner, M., Nuerk, H.-C., Martignon, L., Roesch, S., and Moeller, K. (2016). Finger gnosis predicts a unique but small part of variance in initial arithmetic performance. *J. Exp. Child Psychol.* 146, 1–16. doi: 10.1016/j.jecp.2016.01.006

Weinhold Zulauf, M., Schweiter, M., and von Aster, M. (2003). Das kindergartenalter: sensitive periode für die entwicklung numerischer fertigkeiten. *Kindheit und Entwicklung* 12, 222–230. doi: 10.1026//0942-5403.12.4.222

Weiß, R. H., and Osterland, J. (2013). *Grundintelligenztest Skala 1–Revision (CFT 1-R)*. Göttingen: Hogrefe.

Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., and Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *J. Educ. Psychol.* 102, 43–53. doi: 10.1037/a0016738

Wyschkon, A., Poltz, N., Höse, A., Aster, M. V., and Esser, G. (2015). Schwache fingergnosie als risikofaktor für zukünftiges rechnen? *Lernen und Lernstörungen* 4, 159–175. doi: 10.1024/2235-0977/a000087

Xu, C., and LeFevre, J. (2016). Training young children on sequential relations among numbers and spatial decomposition?: differential transfer to number line and mental transformation tasks. *Dev. Psychol.* 52, 854–866. doi: 10.1037/dev0000124

Conflict of Interest: 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 | Trained conceptual level, skills and games/tasks applied in the training.

Conceptual level	Skill	Game/Task	Order of occurrence in the training. Games were applied with increasing difficulty.	Occurred N-times in training
Finger gnosis (not related to number)	Motoric accuracy	"Finger tapping" with each hands separately and together	1	5
	Motoric divergence	"Labyrinth": tracing a way through a labyrinth with all fingers, separately	2	4
1-to-1 mappings of finger and number	Verbal finger-number mapping	"Naming the fingers," i.e., thumb, index finger, middle finger, ring finger and pinkie and mapping the right numbers (right hand - 1-5 and left hand 6-10)	3	4
	Visual finger-number mapping in association with learning visual Arabic digits	"Tracing numbers" 1-10 on a sheet with the respective finger of the right (1-thumb, 2 - index finger, 3 - middle finger, 4 - ring finger, 5 - pinkie) or left (6-thumb, 7 - index finger, 8 - middle finger, 9 - ring finger, 10 - pinkie) hand	4	3
Ordinal finger-number associations	Counting	"Finger counting" forwards and backwards and starting with different numbers	5	3
		"Counting objects." Children should show the counted objects with their fingers.	12	4
		"Clapping": counting the clapping of the trainer and other children. Children should show the number of claps with their fingers.	13	5
	Ordering (Ordinality based on cardinality)	"Train-Game" with groups of 3-5 children. Each child got an Arabic number. Children had to show their digit with their fingers and order themselves in the correct numerical sequence like train carriages without talking. The number sequences were either continuous with missing "carriages" in between, e.g., 3, 5, and 9	14	6
		"Order card desks" with fingers, digits and points into the right sequence (from 1 to 10)	15	3
Cardinal finger-number associations	Verbal and visual finger to number and number to finger mapping of the respective cardinalities	"Corresponding number" naming the corresponding number to shown fingers (6) The trainer showed a finger-number pattern for a few seconds while saying a rhyme. Children had to recognize the finger-number pattern and show it with their own fingers and say the corresponding number. (7) Later, single children were allowed to show a pattern and appoint another child in solving the task.	6 and 7	12
		"Story-time" detect numbers that were hidden in stories and show the cardinality of the numbers with their fingers"	8	9
		"Memory" with cards displaying fingers and numbers	9	4
		"Bingo" with finger cards Children had to mark numbers on a sheet	10	7
		"Domino" with combined finger and number cards	11	6
Number relations: Base-10 and place-value system	Completion to 5/10	"Completion-Game" showing how many fingers are needed to 5 and 10	16	3
		"Completion" (to 5 and to 10) with cards displaying fingers	17	4
	Double numbers	"Double numbers" children should show the double number (with the fingers) to fingers shown by the trainer	18	3

frontiers in Human Neuroscience

GENERAL COMMENTARY
published: 24 March 2020
doi: 10.3389/fnhum.2020.00099

Commentary: A mental number line in human newborns

Arianna Felisatti[1], Jochen Laubrock[1], Samuel Shaki[2] and Martin H. Fischer[1]*

[1] *Department of Psychology, University of Potsdam, Potsdam, Germany,* [2] *Department of Behavioral Sciences, Ariel University, Ariel, Israel*

Keywords: spatial-numerical associations, SNARC, mental number line (MNL), spatial frequency (SF), temporal frequency, hemispheric asymmetry, newborns, embodied cognition

A Commentary on

A mental number line in human newborns
by Di Giorgio, E., Lunghi, M., Rugani, R., Regolin, L., Dalla Barba, B., Vallortigara, G., et al. (2019). Dev. Sci. 22:e12801. doi: 10.1111/desc.12801

Several thousand (Reuters, 2013) studies have investigated why we associate small numbers with left and large numbers with right space. While humans may learn this association through cultural techniques (Zebian, 2005; Shaki et al., 2009; Göbel et al., 2015), its presence in human new-borns (de Hevia et al., 2017) as well as in non-human animals (Rugani et al., 2015; for review Rugani and de Hevia, 2017; McCrink and de Hevia, 2018) requires a biological explanation. Is there an inborn Spatial-Numerical Association (SNA)?

Di Giorgio et al. (2019) provided a positive answer after testing hour-old humans with a habituation paradigm. They exposed neonates to static two-dimensional images depicting 12 black squares. Once the habituation criterion was reached, bilateral test stimuli were exposed. They consisted of identical images displaying a numerosity which was for some neonates smaller ("4") and for other neonates larger ("36") than the habituated one. Neonates preferred looking at the left image when tested on 4-square images and at the right image when tested on 36-square images.

These findings imply the presence of SNAs at birth; but covariations of numerosity with non-numerical stimulus features prevented clear conclusions. While previously the number of elements was positively correlated with area, a new experiment implemented a negative correlation between numerosity and area by controlling perimeter. Again two groups of new-borns were tested with a single habituation followed by two lateralized test images: Group one after habituating to a 4-big-square image, preferred looking at the 12-small-square image depicted on their right side; conversely group two habituated to a 36-small-square image, preferentially looked at the 12-big-square image displayed on their left side. Since both groups were tested with the same numerosity ("12"), their different looking preferences indicated that they judged the target in relation to the numerosity and not the area of the habituation pattern. The authors interpreted these findings as evidence for an inborn tendency to map numbers onto space, independent of continuous physical variables.

Vallortigara (2018) suggested that few/many elements, triggering withdrawal/approach behaviors, are associated with negative/positive emotions, preferentially processed by the right/left hemisphere, respectively (Davidson, 2004). Instead, we believe that hemispheric specialization for low-level features (Hellige, 1996; Kauffmann et al., 2014) explains the innate SNAs without directly relying on number concepts. Spatial Frequencies (SFs) are defined as number of dark/light cycles/degree of visual angle. Different spatial frequency ranges represent different information

OPEN ACCESS

Edited by:
Reuven Babai,
Tel Aviv University, Israel

Reviewed by:
Giorgio Vallortigara,
University of Trento, Italy

****Correspondence:***
Arianna Felisatti
ariannafelisatti@uni-potsdam.de

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *18 November 2019*
Accepted: *04 March 2020*
Published: *24 March 2020*

Citation:
Felisatti A, Laubrock J, Shaki S and Fischer MH (2020) Commentary: A mental number line in human newborns. Front. Hum. Neurosci. 14:99. doi: 10.3389/fnhum.2020.00099

from any visual scene (Goffaux et al., 2005; Flevaris and Robertson, 2016): Low SFs (few cycles/degree) represent few coarse elements, while high SFs (many cycles/degree) represent many detailed elements. Lateralized vertebrates are neuronally specialized for spatial vision (Vallortigara et al., 2011; Rogers, 2017): Behavioral and neuroscientific studies found that when viewing any scene, vertebrates preferentially extract coarse visual features (low SFs) with their right hemisphere and fine details (high SFs) with their left hemisphere (see **Figure 1A**). This was documented with hierarchical (so-called "Navon") stimuli (Sergent, 1982; Fink et al., 1996); grating/checkboard patterns (see **Figure 1B**; Kitterle and Selig, 1991; Martinez et al., 2001; Piazza and Silver, 2014); and natural scenes (Peyrin et al., 2003; Musel et al., 2013). For any visual scene with homogeneous feature distribution, the cross-over of the optic fibers naturally enhances relative smaller numerosities in our left visual field and relative larger numerosities in our right visual field. In human new-borns, their immature inter-hemispheric communication further augments this bias (Salamy, 1978; Deruelle and de Schonen, 1991).

When we modeled this naturally-occurring visual filtering process on the very stimuli used by Di Giorgio et al. (2019), their behavioral bias emerged from the hemispheric lateralization of SF processing: For any visual scene, patterns with relative few elements preferentially engage the right hemisphere, thus favoring leftward behavior. Conversely, patterns with relative many elements preferentially engage the left hemisphere, thus inducing rightward behavior (**Figure 1C**). Therefore, when total perimeter but not SF content of the stimuli used to test numerosity effects is experimentally controlled, apparent numerical biases reflect natural lateralization of SF processing. If our SF explanation of Di Giorgio et al. (2019) finding is correct, the resulting association "few-left" and "many-right" holds to the degree that numerosity and SF are correlated, e.g., when large numbers tend to be represented by smaller objects. Although our analysis holds for the above habituation study, the same SF filtering principle applies also to viewing stimuli prior to habituation.

More generally, we suggest that our Brain's Asymmetric Frequency Tuning (BAFT) hypothesis accounts for spatial-numerical associations without further need of cognitive mechanisms. Indeed, it provides evidence not only for the origin of horizontal SNAs, but also for their relative nature: Just as the spatial association of small and large numbers depends on the numerical range (Dehaene et al., 1993), the discrimination between low and high SFs depends on the SF range of a given image (Flevaris et al., 2011; Piazza and Silver, 2017).

The BAFT hypothesis makes predictions for numerical cognition and beyond. We predict: (1) In new-borns, for a given numerosity pattern, spatial associations are driven by its absolute or relative SFs; (2) SNAs driven by SFs generalize across cultures and species; (3) SF selection and, as a consequence, SNAs are different in new-borns predisposed to developing autism (enhanced local processing: Jobs et al., 2018) and dyscalculia (deficit in number acuity: Piazza et al., 2010). Moreover, our hypothesis provides a theoretical framework for SNAs across sensory modalities: Indeed, the new-born's association of few syllables with left-space and many syllables with right-space (de Hevia et al., 2017) might reflect temporal frequency tuning in the auditory cortex. The hemispheric asymmetry would be involved in a second stage, after the attentional system has filtered the relevant frequency (double filtering by frequency; Robertson and Ivry, 2000) or could be intrinsic to the process

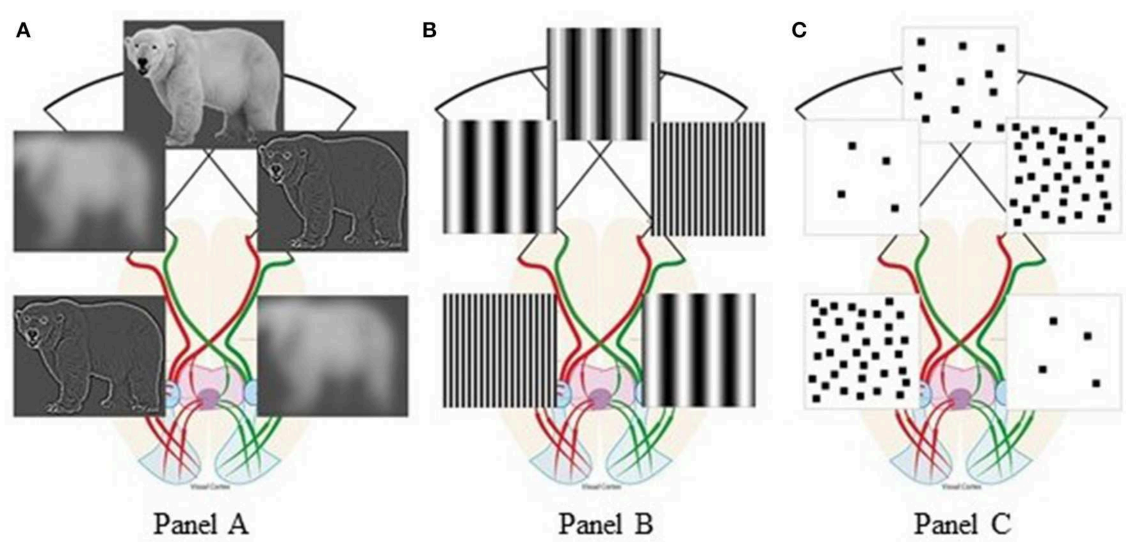


FIGURE 1 | Visual percepts resulting from hemispheric spatial frequency tuning: **(A)** polar bear scenes, adapted from "Figure 3" Panichello et al. (2013), licensed under CC-BY, version 4.0; **(B)** spatial frequency gratings; **(C)** square-pattern stimuli taken from the target article by Di Giorgio et al. (2019). The anatomy of visual pathways is redrawn after "Figure 1. Visual pathway in a primate." by Larsson (2015), used under CC-BY, version 4.0. Figure 1 is licensed under CC-BY, version 4.0 by Arianna Felisatti.

allowing integration of the signal at different temporal windows (asymmetric sampling in time; Poeppel, 2003; Flinker et al., 2019) from early infancy (Telkemeyer et al., 2009).

In conclusion, nature endows us with specialized brains that impose embodied constraints on how we represent numbers.

AUTHOR CONTRIBUTIONS

AF conceived the presented hypothesis. MF and SS contributed to developing the idea. JL helped AF to perform computations on original stimuli used by Di Giorgio et al. (2019) and participated in the discussions. MF supervised all steps of this work. The final commentary reflects the interdisciplinary and international cooperation of AF, JL, SS, and MF.

FUNDING

This work was supported by DFG FI 1915/8-1 Competing heuristics and biases in mental arithmetic and the Open Access Publishing Fund of University of Potsdam.

ACKNOWLEDGMENTS

We thank Elisa Di Giorgio and her colleagues for providing their original materials for our re-analyses.

REFERENCES

Davidson, R. J. (2004). Well–being and affective style: neural substrates and biobehavioural correlates. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 359, 1395–1411. doi: 10.1098/rstb.2004.1510

de Hevia, M. D., Veggiotti, L., Streri, A., and Bonn, C. D. (2017). At birth, humans associate “few” with left and “many” with right. *Curr. Biol.* 27, 3879–3884. doi: 10.1016/j.cub.2017.11.024

Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol. Gen.* 122:371. doi: 10.1037/0096-3445.122.3.371

Deruelle, C., and de Schonen, S. (1991). Hemispheric asymmetries in visual pattern processing in infancy. *Brain Cogn.* 16, 151–179. doi: 10.1016/0278-2626(91)90004-R

Di Giorgio, E., Lunghi, M., Rugani, R., Regolin, L., Dalla Barba, B., Vallortigara, G., et al. (2019). A mental number line in human newborns. *Dev. Sci.* 22:e12801. doi: 10.1111/desc.12801

Fink, G. R., Halligan, P. W., Marshall, J. C., Frith, C. D., Frackowiak, R. S. J., and Dolan, R. J. (1996). Where in the brain does visual attention select the forest and the trees?. *Nature* 382:626. doi: 10.1038/382626a0

Flevaris, A. V., Bentin, S., and Robertson, L. C. (2011). Attentional selection of relative SF mediates global versus local processing: evidence from EEG. *J. Vision* 11:11. doi: 10.1167/11.7.11

Flevaris, A. V., and Robertson, L. C. (2016). Spatial frequency selection and integration of global and local information in visual processing: a selective review and tribute to Shlomo Bentin. *Neuropsychologia* 83, 192–200. doi: 10.1016/j.neuropsychologia.2015.10.024

Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O., and Poeppel, D. (2019). Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nat. Hum. Behav.* 3:393. doi: 10.1038/s41562-019-0548-z

Göbel, S. M., Maier, C. A., and Shaki, S. (2015). Which numbers do you have in mind? Number generation is influenced by reading direction. *Cogn. Process.* 16, 241–244. doi: 10.1007/s10339-015-0715-8

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., and Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception* 34, 77–86. doi: 10.1068/p5370

Hellige, J. B. (1996). Hemispheric asymmetry for visual information processing. *Acta Neurobiol. Exp.* 56, 485–497.

Jobs, E. N., Falck-Ytter, T., and Bölte, S. (2018). Local and global visual processing in 3-year-olds with and without Autism. *J. Autism Dev. Disord.* 48, 2249–2257. doi: 10.1007/s10803-018-3470-8

Kauffmann, L., Ramanoël, S., and Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Front. Integr. Neurosci.* 8:37. doi: 10.3389/fnint.2014.00037

Kitterle, F. L., and Selig, L. M. (1991). Visual field effects in the discrimination of sine-wave gratings. *Percept. Psychophys.* 50, 15–18. doi: 10.3758/BF03212201

Larsson, M. L. (2015). Binocular vision, the optic chiasm, and their associations with vertebrate motor behavior. *Front. Ecol. Evol.* 3:89. doi: 10.3389/fevo.2015.00089

Martinez, A., Di Russo, F., Anllo-Vento, L., and Hillyard, S. A. (2001). Electrophysiological analysis of cortical mechanisms of selective attention to high and low spatial frequencies. *Clin. Neurophysiol.* 112, 1980–1998. doi: 10.1016/S1388-2457(01)00660-5

McCrink, K., and de Hevia, M. D. (2018). From innate spatial biases to enculturated spatial cognition: the case of spatial associations in number and other sequences. *Front. Psychol.* 9:415. doi: 10.3389/fpsyg.2018.00415

Musel, B., Bordier, C., Dojat, M., Pichat, C., Chokron, S., Le Bas, J. F., et al. (2013). Retinotopic and lateralized processing of spatial frequencies in human visual cortex during scene categorization. *J. Cogn. Neurosci.* 25, 1315–1331. doi: 10.1162/jocn_a_00397

Panichello, M. F., Cheung, O. S., and Bar, M. (2013). Predictive feedback and conscious visual experience. *Front. Psychol.* 3:620. doi: 10.3389/fpsyg.2012.00620

Peyrin, C., Chauvin, A., Chokron, S., and Marendaz, C. (2003). Hemispheric specialization for spatial frequency processing in the analysis of natural scenes. *Brain Cogn.* 53, 278–282. doi: 10.1016/S0278-2626(03)00126-X

Piazza, E. A., and Silver, M. A. (2014). Persistent hemispheric differences in the perceptual selection of spatial frequencies. *J. Cogn. Neurosci.* 26, 2021–2027. doi: 10.1162/jocn_a_00606

Piazza, E. A., and Silver, M. A. (2017). Relative spatial frequency processing drives hemispheric asymmetry in conscious awareness. *Front. Psychol.* 8:559. doi: 10.3389/fpsyg.2017.00559

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

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Commun.* 41, 245–255. doi: 10.1016/S0167-6393(02)00107-3

Reuters, T. (2013). *Web of Science 2010*. Available online at: http://thomsonreuters.com/products_services/science/science_products/az/web_of_science/ (accessed).

Robertson, L. C., and Ivry, R. (2000). Hemispheric asymmetries: attention to visual and auditory primitives. *Curr. Dir. Psychol. Sci.* 9, 59–63. doi: 10.1111/1467-8721.00061

Rogers, L. J. (2017). A matter of degree: strength of brain asymmetry and behaviour. *Symmetry* 9:57. doi: 10.3390/sym9040057

Rugani, R., and de Hevia, M. D. (2017). Number-space associations without language: evidence from preverbal human infants and non-human animal species. *Psychon. Bull. Rev.* 24, 352–369. doi: 10.3758/s13423-016-1126-2

Rugani, R., Vallortigara, G., Priftis, K., and Regolin, L. (2015). Number-space mapping in the newborn chick resembles humans’ mental number line. *Science* 347, 534–536. doi: 10.1126/science.aaa1379

Salamy, A. (1978). Commissural transmission: maturational changes in humans. *Science* 200, 1409–1411. doi: 10.1126/science.208144

Sergent, J. (1982). The cerebral balance of power: confrontation or cooperation? *J. Exp. Psychol. Hum. Percept. Perform.* 8:253. doi: 10.1037/0096-1523.8.2.253

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

Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., et al. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *J. Neurosci.* 29, 14726–14733. doi: 10.1523/JNEUROSCI.1246-09.2009

Vallortigara, G. (2018). Comparative cognition of number and space: the case of geometry and of the mental number line. *Philos. Trans. R. Soc. B Biol. Sci.* 373:20170120. doi: 10.1098/rstb.2017.0120

Vallortigara, G., Chiandetti, C., and Sovrano, V. A. (2011). Brain asymmetry (animal). *Wiley Interdiscipl. Rev. Cogn. Sci.* 2, 146–157. doi: 10.1002/wcs.100

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

Conflict of Interest: 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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 24 March 2020
doi: 10.3389/fnhum.2020.00088

Hemispheric Lateralization of Arithmetic Facts and Magnitude Processing for Two-Digit Numbers

Stefanie Jung[1,2], Korbinian Moeller[1,2,3], Hans-Otto Karnath[4] and Elise Klein[1,2,5]*

[1] Junior Research Group Neuro-Cognitive Plasticity, Leibniz-Institut für Wissensmedien, Tübingen, Germany, [2] Research Methods and Mathematical Psychology, Eberhard Karls Universität Tübingen, Tübingen, Germany, [3] LEAD Graduate School & Research Network, University of Tübingen, Tübingen, Germany, [4] Center of Neurology, Section for Neuropsychology, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany, [5] CNRS UMR 8240, Laboratory for the Psychology of Child Development and Education, Paris, France

In the human brain, a (relative) functional asymmetry (i.e., laterality; functional and performance differences between the two cerebral hemispheres) exists for a variety of cognitive domains (e.g., language, visual-spatial processing, etc.). For numerical cognition, both bi-lateral and unilateral processing has been proposed with the retrieval of arithmetic facts postulated as being lateralized to the left hemisphere. In this study, we aimed at evaluating this claim by investigating whether processing of multiplicatively related triplets in a number bisection task (e.g., 12_16_20) in healthy participants (n = 23) shows a significant advantage when transmitted to the right hemisphere only as compared to transmission to the left hemisphere. As expected, a control task revealed that stimulus presentation to the left or both visual hemifields did not increase processing disadvantages of unit-decade incompatible number pairs in magnitude comparison. For the number bisection task, we replicated the multiplicativity effect. However, in contrast to the hypothesis deriving from the triple code model, we did not observe significant hemispheric processing asymmetries for multiplicative items. We suggest that participants resorted to keep number triplets in verbal working memory after perceiving them only very briefly for 150 ms. Rehearsal of the three numbers was probably slow and time-consuming so allowing for interhemispheric communication in the meantime. We suggest that an effect of lateralized presentation may only be expected for early effects when the task is sufficiently easy.

Keywords: interhemispheric communication, number comparison task, number bisection task, two-digit number processing, hemispheric lateralization

OPEN ACCESS

Edited by:
Carmen Moret-Tatay,
Valencia Catholic University Saint Vincent Martyr, Spain

Reviewed by:
Sebastian Ocklenburg,
Ruhr-University Bochum, Germany
Tommaso Bocci,
University of Milan, Italy

****Correspondence:***
Stefanie Jung
s.jung@iwm-tuebingen.de

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *18 November 2019*
Accepted: *27 February 2020*
Published: *24 March 2020*

Citation:
Jung S, Moeller K, Karnath H-O and Klein E (2020) Hemispheric Lateralization of Arithmetic Facts and Magnitude Processing for Two-Digit Numbers. Front. Hum. Neurosci. 14:88. doi: 10.3389/fnhum.2020.00088

INTRODUCTION

One of the most important postulates of the Triple Code Model (henceforth TCM) of numerical cognition is the distinction between the representation of number magnitude processing on the one hand and arithmetic facts and their verbally mediated retrieval from long term memory on the other hand (Dehaene and Cohen, 1995, 1997; Dehaene et al., 2003). As regards number magnitude processing, the TCM suggests a bilateral fronto-parietal network around the intraparietal sulcus (IPS) to be dedicated to the representation and mental manipulation of numerical quantities – for instance, when calculations need to be performed (e.g., 124–56). In contrast, tasks such as multiplication with small numbers (e.g., 3 × 2) are supposed to be solved by arithmetic fact retrieval subserved by a left-hemispheric network including perisylvian language areas as well as the angular gyrus (Dehaene et al., 2003). As such number magnitude is assumed to be represented redundantly in both hemispheres of the human brain, whereas the verbal representation of arithmetic facts is postulated for the left hemisphere of the human brain only.

The current view is that arithmetic facts are stored and retrieved in a verbal code (Dehaene et al., 2003). Neuro-functional evidence on the neural networks underlying verbally mediated fact retrieval stems primarily from studies that investigating the acquisition of arithmetic facts by means of drill trainings of difficult multiplication problems (e.g., 43 × 9 = ___; Bloechle et al., 2016; Delazer et al., 2003, 2005; Grabner et al., 2009; Ischebeck et al., 2006). Consistently, stronger activation was found in left-hemispheric perisylvian language areas as well as the left angular gyrus (e.g., Delazer et al., 2003, 2005; Ischebeck et al., 2006; Grabner et al., 2009) and the left hippocampus (e.g., Bloechle et al., 2016; Klein et al., 2019) for trained problems as opposed to untrained problems after the training. These activation patterns are assumed to reflect automatic verbally mediated retrieval of arithmetic facts from long-term memory (Delazer et al., 2003; Ischebeck et al., 2006; Bloechle et al., 2016).

In order to investigate the processing of arithmetic facts and number magnitude within one task, the number bisection task (NBT; Nuerk et al., 2002) was proposed. In the NBT, participants have to evaluate whether the central number of a triplet (e.g., 11_13_15) corresponds to the arithmetic integer mean of the interval defined by the two outer numbers. Triplets which are part of a multiplication table (21_24_27) provided a processing advantage as compared to non-multiplicative triplets (19_22_25, cf. Nuerk et al., 2002) by activating multiplication fact knowledge. Wood et al. (2008) replicated these findings and observed that processing of multiplicative triplets was specifically associated with activation in left-hemispheric perisylvian language areas and the angular gyrus (see also Klein et al., 2016 for a re-analysis). However, concurrent articulation led to relative slowing of processing multiplicative triplets in the NBT, which reduced the multiplicativity effect (Moeller et al., 2011).

These results support the central postulation of the TCM that arithmetic facts are processed in the left hemisphere only. This argument is primarily based on classical neuropsychological single-case studies on brain-lesioned patients. For instance, patient BOO (Dehaene and Cohen, 1997), patient WT (Zaunmuller et al., 2009) or patient VOL (Cohen and Dehaene, 2000), who suffered from left-hemispheric lesions, showed severe selective impairments in multiplication, which is solved by arithmetic fact retrieval. However, existing case studies of brain-lesioned patients do not support the assumption of the TCM that arithmetic facts are processed in a left-lateralized manner consistently: for instance, Granà et al. (2006) reported the case of patient PN who showed circumscribed deficits in multiplication following a right-hemispheric lesion. Moreover, Salillas and coworkers (Salillas et al., 2012) reported an association of multiplication performance and the right IPS by inducing a virtual lesion using transcranial magnetic stimulation.

In view of these inconsistent findings, the question whether the verbal representation of arithmetic facts is indeed lateralized to the left hemisphere is far from being answered comprehensively. Also, findings from various fMRI studies cannot provide sufficient evidence for isolated left-hemispheric activation for arithmetic fact retrieval as they often observed bilateral activation of perisylvian language areas (e.g., Dehaene et al., 1999; Stanescu-Cosson et al., 2000; Grabner et al., 2009; Klein et al., 2013a,b, 2016; Bloechle et al., 2016).

Therefore, it would be important to obtain converging evidence from healthy adult participants substantiating the theoretical claim that verbally mediated arithmetic fact retrieval is lateralized to the left hemisphere.

To this end, we used a task indicative of arithmetic fact retrieval in a divided visual field paradigm. In this divided visual field paradigm, respective stimuli are presented either unilaterally in the right or the left visual hemifield or bilaterally in both visual fields. When the stimuli are presented unilaterally into one visual hemifield only, visual input is initially only transmitted into the contralateral hemisphere. Evidence for the successful application of such divided visual field paradigms can be found in various domains (e.g., language: Geffen et al., 1971; Brysbaert, 1994; numerical cognition: Dimond and Beaumont, 1971; Hatta et al., 2002; Ratinckx et al., 2006; Hildebrandt et al., 2016). To give an example for the principle of this paradigm, when a stimulus is presented in the left visual hemifield, it would first be transferred to the right hemisphere; left-hemispheric processing of the respective stimuli would only occur after further transmission of the processed stimulus to the left hemisphere via interhemispheric transcallosal fibers. In case arithmetic facts are indeed processed exclusively in the left hemisphere, unilateral input into the left visual hemifield and thus initial transmission to the right hemisphere should lead to a processing disadvantage, reflected by, for instance, longer response latencies and lower accuracy compared to unilateral presentation of the stimuli into the right visual hemifield. Evidence for interhemispheric processing and its modulation has been provided by several studies using tDCS on the non-dominant hemisphere for the task at hand in higher cognitive processing, such as arithmetic fact retrieval (e.g., Clemens et al., 2013; for anodal stimulation) and primary perceptual processing (e.g., Bocci et al., 2018; for vision acuity).

Ratinckx et al. (2006) demonstrated that the divided visual field paradigm showed differential effects for the case of the unit-decade compatibility effect in two-digit number magnitude comparison, which is supposed to be processed in the right hemisphere (Wood et al., 2006; for a review see Nuerk et al., 2011). The unit-decade compatibility effect describes the finding that magnitude comparison in compatible number pairs (i.e., when separate comparisons of decade and unit digits lead to the same decision, e.g., in 42_57, 4 < 5 and 2 < 7) is easier than in incompatible number pairs (in which separate comparisons of decade and unit digits lead to opposing decisions, e.g., in 47_62; 4 < 6, but 7 > 2). In the study by Ratinckx et al. (2006), the disadvantage for the more demanding incompatible items was smaller when stimuli were presented unilaterally in the left visual hemifield or bilaterally to both visual fields and thus allowed initial right-hemispheric processing.

In the current study, we aimed at realizing a similar setting for the retrieval of arithmetic facts. To this end, we evaluated modulations of the multiplicativity effect in the NBT (Nuerk et al., 2002; Wood et al., 2008) in a divided visual field paradigm. In the NBT, the multiplicativity effect describes faster response times and lower error rates for triplets, which are part of a multiplication table (e.g., 21_24_27) as compared to number triplets which are not (e.g., 22_25_28, Nuerk et al., 2002; Moeller et al., 2009). Additionally, the multiplicativity effect was associated with activation of left-hemispheric language areas and the angular gyrus (Wood et al., 2008). It has been argued that multiplicativity of a triplet provides a processing advantage by activating multiplication fact knowledge (Nuerk et al., 2002).

For our divided visual field paradigm on the NBT, we used the stimulus set of Moeller et al. (2009). As a control task, we replicated the experiment by Ratinckx et al. (2006) on the unit-decade compatibility effect in magnitude comparison. To ensure that results are not confounded by stimulus specificities we created magnitude comparison stimuli only using numbers from the NBT stimulus set. This way, the magnitude comparison task served two purposes: on the one hand, it was used to verify that participants indeed could perceive and process the respective two-digit numbers which were presented only briefly at perifoveal positions. On the other hand, the task was used as a proof of concept: by replicating the results of Ratinckx et al. (2006) on modulation of the compatibility effect by lateralized presentation, we aimed at verifying that our experimental setting was valid.

In sum, the present study aimed at investigating whether the verbal representation of arithmetic facts is indeed lateralized to the left hemisphere of the human brain as put forward by the TCM (e.g., Dehaene and Cohen, 1995, 1997; Dehaene et al., 2003). Therefore, we investigated whether the processing of multiplicative triplets shows a significant advantage when stimuli are initially transmitted to the right hemisphere only. In particular, our hypotheses were as follows: as regards the number magnitude comparison control task, we expected to replicate the results of Ratinckx et al. (2006) of a modulation of the compatibility effect by lateralized presentation. In particular, presentation of number pairs in the left or bilaterally in both visual hemifields should reduce the disadvantage for incompatible number pairs in magnitude comparison. With respect to the NBT, we expected to replicate the multiplicativity effect. However, multiplicativity should only facilitate bisection performance when items were presented in the right visual hemifield or bilaterally in both visual fields because in this case input is directly transmitted to the left hemisphere of the brain for which the verbal representation of arithmetic facts is postulated.

MATERIALS AND METHODS

Participants

Prior to data collection, we calculated the necessary sample size for the used within-participant design comparing effects of lateralized stimulus presentation for magnitude comparison and arithmetic fact retrieval in the number bisection task based on effect sizes reported in prior studies. For the multiplicativity effect in the NBT, both small (Cohen's d = 0.2–0.4, Moeller et al., 2011) and large effect sizes (Cohen's d = 0.8, Nuerk et al., 2002) were observed so far. For the effect of lateralization of presentation, a medium effect size was found the study by Ratinckx et al. (2006). Expecting a medium effect size of about d = 0.6 for both effects, a sample size of 21 participants should allow for detecting the respective effect with enough statistical power. In particular, we used the following parameters for the *a priori* sample calculation: As we used a repeated measure within-participant design, we considered one group of participants. For the effect size, we assumed a partial eta square of $\eta^2_p = 0.20$. We expected an alpha error probability of $p = 0.05$ and a power of 0.95. Furthermore, we compared three different measurements (i.e., bilateral, right lateralized and left-lateralized item presentation). Among these repeated measures, we expected a high correlation of 0.85.

In total, 32 right-handed healthy volunteers (7 male, mean age = 24.5 years; SD = 3.56), who graded '4' or better in mathematics in their school-leaving certificate (with grades in Germany ranging from 1 to 6 with 1 being the grade), were recruited via public announcements. All participants had normal or corrected to normal vision and reported no history of neurological or psychiatric disorders. We excluded one participant from data analysis as she reported to have suffered from math anxiety during school.

Thus, data from 31 participants (6 male, mean age = 24.34 years; SD = 3.03), were considered for the analyses. 30 participants had more than 10 years of formal education. Eight participants are exposed to mathematics in their profession. The Edinburgh Handedness Inventory (Oldfield, 1971) was used to determine handedness. Participants were categorized as right—handed using the cut-off criterion of LQ $> +50$, indicating them to fall in the first decile or higher of right-handedness. Eye dominance was also recorded for both distance (right = 21 participants, left = 10 participants) and proximity (right = 11 participants, left = 3 participants, not defined = 17 participants).

The study was approved by the local Ethics Committee (082/2018BO2) and was performed in compliance with the latest version of the Code of Ethics of the World Medical Association (Declaration of Helsinki). All participants gave their written informed consent prior to the study and received compensatory payment.

Procedure

Data were collected in individual 2-h testing sessions. Within one session, two experimental tasks had to be completed: a number magnitude comparison task and a NBT. The order of both tasks was counterbalanced across participants to minimize order effects. These two experimental tasks were followed by a control task assessing multiplication fact retrieval. Task instructions emphasized both speed and accuracy in all tasks. Furthermore, the left and right control key on the keyboard were used as response buttons in all tasks. Stimulus presentation, response times and accuracy were recorded using Presentation software version 20.03 (Neurobehavioral Systems Inc., Albany, CA, United States).

Task and Stimuli

Both in the experimental and the control task participants were required to give "Yes" or "No" responses by pressing either the left (i.e., "No" response: left Ctrl-key press) or the right response button (i.e., "Yes" response: right Ctrl-key press) with their left and right index finger, respectively. In all tasks, problems were presented in pseudo-randomized order, preventing a direct repetition of the same problem. Additionally, the sequence was manipulated such that no more than three correct or false trials, respectively, were presented in a row. This also applied to the side of presentation (i.e., left, right, bilateral) of items in both experimental tasks.

Number Bisection Task

In the NBT, 200 two-digit number triplets (100 correctly bisected: e.g., 18_24_30 and 100 incorrectly bisected: e.g., 17_18_30), covering the range from 11 to 99 were presented (cf. Moeller et al., 2009 for the same item set). The same item set was used for each condition (i.e., right, left, and bilateral), resulting in a total of 600 trials. At the beginning, twelve randomly chosen triplets were used as practice trials.

We used a 2 × 3 design for correctly bisected triplets (requiring "Yes" responses) as well as incorrectly bisected triplets (requiring "No" responses). For correctly bisected triplets, the factors *multiplicativity* (yes: e.g., 21_24_27 vs. no: 22_25_28) and *lateralization* (i.e., right vs. left vs. bilateral) were manipulated. For incorrectly bisected triplets, *bisection possibility* (bisectable: e.g., 21_22_27 vs. non-bisectable: e.g., 23_26_30) and again *lateralization* (i.e., right vs. left vs. bilateral) was varied. A comprehensive description of stimulus can be found in Moeller et al., 2009).

Participants had to decide whether the triplet's central number represented the arithmetic mean of the two outer numbers. They were required to indicate their decision by pressing either the left (i.e., "No response") or the right response button (i.e., "Yes" response). The experiment allowed participants to take a self-defined break after 50 trials each.

Number Comparison Task

For the number comparison task, a subset of triplets from the study by Moeller et al. (2009) was used, whereby only the two outer numbers were offered as duplets to be compared (e.g., 22_38). This subset included 75 item pairs,[1] 25 within-decade items with two numbers within the same decade (e.g., 22_28). The remaining 50 items were manipulated for unit-decade compatibility (i.e., compatible vs. incompatible trials). As Ratinckx et al. (2006) only observed an effect of lateralized stimuli presentation on the compatibility of the presented number pairs, we focused on the manipulation of this factor to reduce the number of items and thus total testing time. The same item set was used for each condition (i.e., right, left, and bilateral), resulting in a total of 225 number comparison tasks. All items were presented in pseudo-randomized order in one run. The same practice trials were used as in the NBT but without presentation of the central number of the triplet. Participants had to decide whether the upper number was larger than the lower number on the display by pressing either the left (i.e., "No" response) or the right response button (i.e., "Yes" response). **Appendix Table A1** provides an overview of the stimulus set for the number magnitude comparison tasks.

Control Task

In the control task, participants' multiplication performance was assessed to consider it in the subsequent analyses. Therefore, we used the same verification paradigm and experimental setup as used by Clemens et al. (2013). One hundred and eighty simple multiplication problems (90 with a correct and 90 with an incorrect solution probe), covering the operand range from 0 to 10, were presented in Arabic format (e.g., 7 × 5 = 35). Multiplication problems included standard problems (96), rule problems (72, with 24 problems using 0, 1, and 10 as multiplicand, respectively), and tie problems (12). In addition, 30 different multiplication problems (15 with a correct and 15 with an incorrect solution probe) were used as practice trials. Participants had to decide whether the presented solution probe of the multiplication problem was correct (i.e., "Yes" response) or incorrect (i.e., "No"- response). **Appendix Table A2** in the **Appendix** provides an overview of the stimulus set for the control tasks.

Apparatus, Experimental Paradigms, and Stimulus Presentation

For all experiments, a 17″ screen driven at a resolution of 1920 × 1200 pixels and 60 Hz refreshing rate was used. Participants were seated at a distance of 60 cm from the screen. Constant viewing distance was ensured by using a head and chin rest. During the experiment, the experimenter was sitting directly opposite to the participant to control eye fixation. In case of a loss of fixation, the experimenter reminded the participant to fixate the center of the screen, which was only necessary in a few trials in 6 of 23 participants. Participants gave their answers on a standard QWERTZ keyboard with the keys necessary for the experiments labeled. In-ear headphones for the examiner were connected to the laptop so that an acoustic signal, which only the examiner could hear, could be used to indicate the beginning of each trial.

In the divided visual field paradigm, experimental setup and stimulus presentation for the experimental tasks (i.e., number comparison and NBT) was similar to the study of Ratinckx et al. (2006). Items were presented tachistoscopically in four different visual hemifield displays, this means, two unilateral conditions (right and left) and two bilateral conditions. For the bilateral conditions, the item set was split in half so that numbers in the first half of items were displayed in the upper left and lower right corner, respectively (i.e., bilateral A), and numbers in the second half of items were presented in the lower left and upper right corner (i.e., bilateral-B). Bilateral condition A and B were collapsed for subsequent data analysis. In the experimental design of Ratinckx et al. (2006), however, visual input differed between

[1] Out of the 75 item pairs, 74 are taken from the triplets by Moeller et al. (2009) and one additional item was generated.

the unilateral and bilateral condition as follows: In the unilateral condition, target items were presented either to the left or to the right of a centrally presented fixation cross. In the bilateral condition, target items were presented to the left and right of the fixation cross while empty positions were filled "##."

In the present study, we aimed at balancing perceptual load across unilateral and bilateral presentation of stimuli. Thefore, we adapted the experimental paradigm slightly (cf. **Figure 1**): In the center of the screen, a fixation cross (i.e., plus [+] sign, font: Arial, size: 30), which extended 0.5° of visual angle horizontally and vertically, was presented. An imaginary square measuring 5 °–5 ° (cf. Ratinckx et al., 2006; for the trigonometrical constraints of the bilateral condition) was centered around the fixation stimulus. The diagonal projection between fixation location and corners of the imaginary square followed a 45° angle.

Two-digit Arabic numbers ranging from 11 to 99 were presented (extending 1.7° of visual angle horizontally and 1.2° vertically) at the corners of the imaginary square centered around the fixation cross. At corners not occupied by numerical stimuli, "##" was presented to keep visual input comparable across conditions (see **Figure 2**). For example, in the right unilateral presentation condition, numbers were displayed to the right of the fixation cross and "##" were displayed left from the fixation cross. **Figure 1** provides an overview of the experimental setup. Due to this setup, conditions were presented in a randomized order (cf. Ratinckx et al., 2006, for conditions presented block-wise).

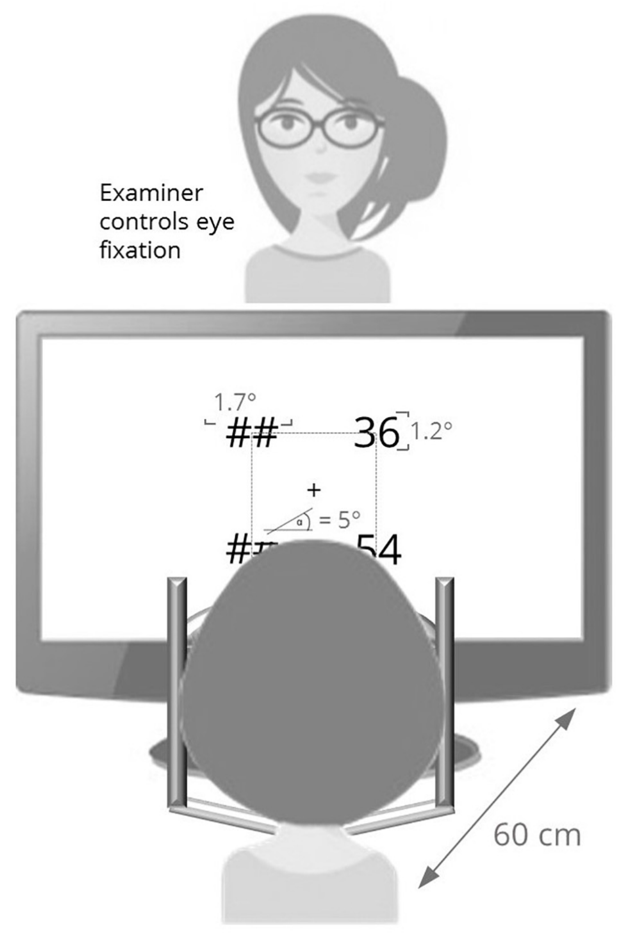


FIGURE 1 | Experimental setup for the lateralized stimulus presentation.

At the beginning of each trial, the fixation cross was displayed for 600 ms. Simultaneously, an acoustic signal was presented to the experimenter through the headphones. This acoustic signal indicated the beginning of a new trial to the experimenter because he/she was unable to see the screen from his/her position but had to monitor eye-movements of the participants. In the case of the NBT, the central number of the upcoming triplet was presented at the position of the fixation cross for another 600 ms. Then, "##" replaced the central number and the triplet's outer numbers were tachistoscopically presented for 150 ms at two corners of the imaginary square. Finally, all positions were covered by "##" for a maximum of 3650 ms or until a response key was pressed. In total, one trial lasted up to 5000 ms. **Figure 2** illustrates the trial sequencing. In experimental trials, participants were instructed to fixate the fixation sign throughout the entire trial and not to move their eyes.

In the control task, items were not presented in a divided field paradigm. Instead, we used the same experimental setup as Clemens et al. (2013). In this setup, the overall presentation time for each multiplication problem was variable. The multiplication problem disappeared immediately after the response was given with a maximum presentation time of 3000 ms. Each multiplication problem was followed by a mask ("######") presented for 500 ms, to keep trials separated from each other.

Data Analysis

Data were analyzed using the open source language and statistical environment R (Version 3.6.; R Core Team, 2019). All analyses were done on the rate correct score (RCS; Woltz and Was, 2006), a combined speed (RT) and accuracy measure of performance. The RCS was calculated by combining the proportion of correctly solved trials and average RT for each condition (i.e., lateralization, multiplicativity, compatibility, etc.) to make the measure comparable across conditions (cf. Vandierendonck, 2017) to reflect the number of correct answers per second. Participants were excluded from data analysis when they scored less than 60% correct in one of the tasks. Only RTs for correct responses (both "Yes" and "No" answers) larger than 200 ms in the number bisection task and 150 ms in the number comparison task were analyzed. Incorrect or missing responses were not considered.

In the NBT, effects of multiplicativity, bisection possibility and lateralization on the RCS were evaluated separately for correctly bisected and incorrectly bisected triplets. Additionally, the effect of multiplicativity was controlled for individual multiplication performance assessed by the control task. In the number comparison task, we analyzed influences of compatibility and lateralization on the RCS. Prior to this, we compared participants' performance (i.e., mean percentage correct and mean RT) in our study to participants' performance in the study by Ratinckx et al. (2006).

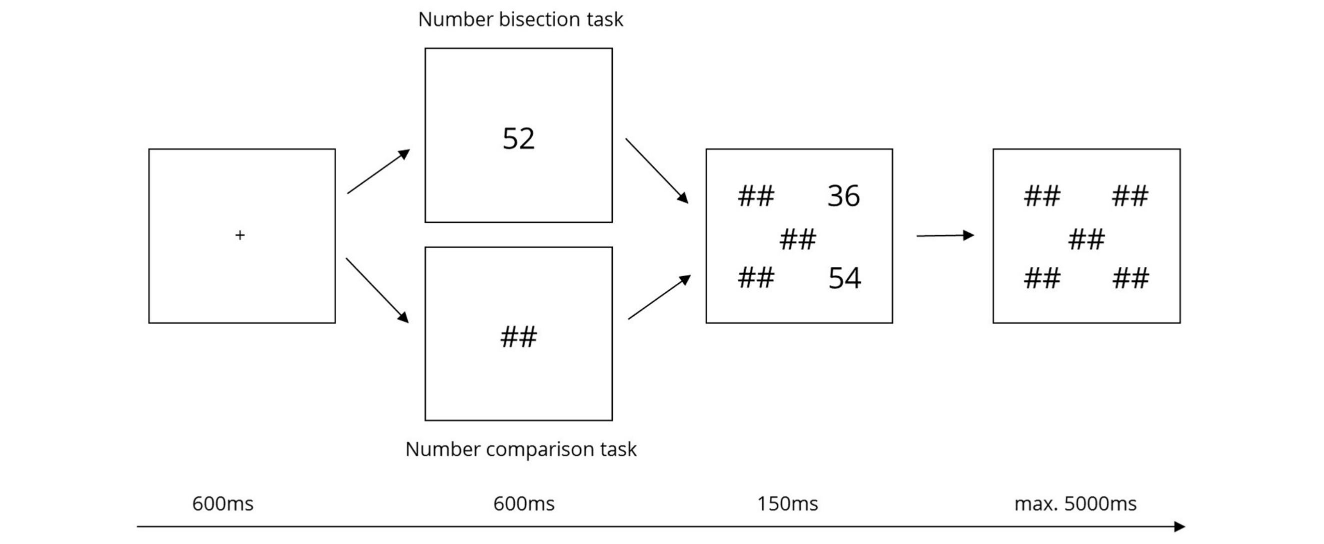


FIGURE 2 | Trial sequence for the NBT (top) and number comparison task (bottom).

RESULTS

In total, complete data sets of 23 participants entered analyses (6 male, mean age: 24.34, *SD* = 3.03). Eight participants had to be excluded for scoring below 60% correct in the NBT. All of these participants had more than 10 years of formal education. Mean LQ was 84.53 (*SD* = 16.01) according to the Edinburgh handedness inventory. Eye dominance was defined as predominantly right for the distance (right = 14 participants, left = 9 participants). **Table 1** summarizes the results of the two experimental tasks and the control task (i.e., percentage correct, reactions times, and the RCS).

Number Bisection Task

First, correlating reaction times and percentage correct trials revealed a relatively high correlation (Spearman correlation: r_s = 0.49). This correlation suggests the presence of a speed-accuracy trade-off in solving the NBT and, thus, may warrant the use of the RCS in subsequent analyses.

For correctly bisected triplets, we evaluated modulations of the multiplicativity effect using a 2 (multiplication: multiplicative triples vs. non-multiplicative triples) × 3 (lateralization: right vs. left vs. bilateral) analysis of covariance (ANCOVA) with the RCS as dependent variable. There was a significant main effect of multiplicativity [$F(1,131) = 47.08$, $p < 0.001$, $\eta^2_p = 0.23$)] prevailing after controlling for test performance in multiplication facts [$F(1,131) = 29.79$, $p < 0.001$, $\eta^2_p = 0.14$)]. This indicated that triplets which are part of multiplication tables were responded to with more correct responses per second as compared to triplets not part of multiplication tables (see **Figure 3A**). The main effect for lateralization was not significant [$F(2,131) = 0.19$, $p = 0.82$, $\eta^2_p = 0.0018$], neither was the interaction [$F(2,131) = 0.058$, $p = 0.94$, $\eta^2_p < 0.001$].

Additionally, we conducted a *post hoc* analysis in order to further investigate the null effect of lateralization on multiplicative items in the NBT. To this end, we included the problem size in our *post hoc* analysis by conducting a median split. We considered large problem sizes to be those where the smallest of a triplet's numbers was greater than 46. When our items were less reflective of fact knowledge and required more cognitive demand, multiplicativity should play a smaller role in larger problem sizes. In other words, triplets like 18_27_36 would be more closely associated with being dividable by 9 than triplets like 63_72_81. Therefore, we ran a multiple linear regression predicting RCS based on multiplicativity and problem size of items to check this hypothesis.

A significant regression equation was found with an R^2 of 0.66 [$F(3, 260) = 172.2$, $p < 0.001$]. Multiplicativity significantly increased participant's RCS. Problem size instead was not a significant predictor of the RCS. However, the interaction of multiplicativity and problem size was significant with $p < 0.001$, indicating that multiplicativity of triplets was specifically beneficial when triplets consisted of smaller numbers. Results are displayed in **Table 2**.

For incorrectly bisected triplets, we evaluated the effect of bisection possibility and lateralization using a 2 (bisection possibility: bisectable vs. non-bisectable triples) × 3 (lateralization: right vs. left vs. bilateral) analysis of variance (ANOVA) with the RCS as the dependent variable. There was no significant main effect for bisection possibility [$F(1,132) = 2.16$, $p = 0.14$, $\eta^2_p = 0.002$], indicating that task efficiency did not differ between bisectable (M = 0.45, SD = 0.02) and non-bisectable triplets (M = 0.48, SD = 0.11). Moreover, there was no main effect for lateralization [$F(2,132) = 0.87$, $p = 0.42$, $\eta^2_p = 0.012$] nor a significant interaction [$F(2,132) = 0.36$, $p < 0.69$, $\eta^2_p = 0.005$].

Number Comparison Task

First, in view of the rather poor performance of participants in the NBT, we checked whether participants' performance (i.e., mean percentage correct and mean RT) in the number comparison task

TABLE 1 | Overview of the mean test performance in the different tasks providing mean percentage correct sores, mean reaction times and the mean rate correct score (RCS) for the different tasks.

	Right		Left		Bilateral	
	Mean	***SD***	**Mean**	***SD***	**Mean**	***SD***
Number bisection task						
Multiplicative triplets						
Percentage correct	61.56	13.90	62.00	13.65	64.17	13.33
Reaction time	1666.52	447.04	1668.87	443.43	1758.54	467.15
Rate correct score (RCS)	0.40	0.15	0.41	0.16	0.39	0.15
Non-multiplicative triplets						
Percentage correct	37.39	7.22	39.96	7.08	37.13	5.52
Reaction time	1470.50	385.38	1515.72	407.71	1566.29	405.71
Rate correct score (RCS)	0.28	0.10	0.26	0.09	0.24	0.08
Bisectable triplets						
Percentage correct	72.87	9.99	66.87	10.82	69.72	11.66
Reaction time	1607.36	420.62	1603.21	381.89	1695.61	374.00
Rate correct score (RCS)	0.48	0.14	0.43	0.12	0.43	0.12
Non-bisectable triplets						
Percentage correct	75.72	11.04	72.60	12.68	76.60	11.64
Reaction time	1614.37	392.82	1594.32	396.56	1665.54	354.88
Rate correct score (RCS)	0.49	0.11	0.48	0.14	0.48	0.12
Number bisection task						
Compatible pairs						
Percentage correct	85.73	5.63	83.82	5.04	84.87	8.44
Reaction time	919.49	170.65	921.00	200.65	919.10	191.94
Rate correct score (RCS)	1.20	0.39	1.00	0.22	0.90	0.17
Incompatible pairs						
Percentage correct	77.04	12.42	81.91	12.17	83.82	11.78
Reaction time	908.74	173.48	910.33	204.68	944.12	167.68
Rate correct score (RCS)	0.89	0.21	0.74	0.15	1.04	0.32
Control task						
Percentage correct	–	–	–	–	90.50	4.65
Reaction time	–	–	–	–	1248.09	215.8
Rate correct score (RCS)	–	–	–	–	0.75	0.15

in the current study was comparable to participants' performance reported by Ratinckx et al. (2006). **Figures 4A,B** illustrates participants' performance in both studies.

Independent-samples t-tests were conducted, one for each lateralization condition (i.e., right, left, bilateral), separately for correctness and reaction times. p-Values were corrected for multiple applying the procedure suggested by Bonferroni. In terms of correctness, significant differences were observed between lateralization conditions in the present study (right: M = 82.31, SD = 6.68, left: M = 83.71, SD = 6.53, bilateral: M = 83.82, SD = 7.92) and the study by Ratinckx et al. (2006; right: M = 87, left: M = 88, bilateral: M = 90), $t_{min}(22) = 3.37$, $p < 0.001$. Despite the significant differences, performance differed by no more than seven percentage errors. Interestingly, correctness in both studies was highest in the bilateral condition and lowest in the right lateralized condition.

In terms of reaction times, participants in the current study showed on average longer reaction times (right: M = 896 ms, SD = 171, left: M = 912 ms, SD = 189, bilateral: M = 924 ms, SD = 178.46) than did participants reported by Ratinckx and colleagues (right: M = 658 ms, left: M = 656 ms, bilateral: M = 637 ms), $t_{min}(22) = 7.73$, $p < 0.001$ – suggesting higher task demands in the current study.

Second, influences of lateralized presentation of stimuli on the unit-decade compatibility effect were evaluated by running 2 (compatibility: compatible vs. incompatible pairs) × 3 (lateralization: right vs. left vs. bilateral) ANOVA on the RCS as the dependent variable. There was a marginally significant main effect of compatibility [$F(1,132) = 3.51$, $p = 0.06$, $\eta^2_p = 0.02$], indicating more correctly solved items per second for compatible (M = 1.06, SD = 0.28) as compared to incompatible number pairs (M = 0.93, SD = 0.27). The main effect for lateralization was not significant [$F(2,132) = 0.46$, $p = 0.66$, $\eta^2_p = 0.0004$]. However, the interaction of compatibility and lateralization was significant [$F(2,132) = 17.78$, $p < 0.001$, $\eta^2_p = 0.21$, illustrated in **Figure 3B**]. The same interaction was found in the study by Ratinckx et al. (2006).

Post hoc comparisons for lateralization using Tukey HSD controlling for multiple comparisons showed a significant [$p = 0.002$] difference between presentation to the right visual hemifield (M = 1.20, SD = 0.39) and bilateral presentation (M = 0.90, SD = 0.17) for compatible number pairs. Differences between these two conditions and presentation to the left visual hemifield (M = 1.00, SD = 0.22) were not significant. For incompatible number pairs, significant differences were observed between presentation to the right hemifield condition (M = 0.74, SD = 0.14) and both presentation to the left hemifield (M = 1.04, SD = 0.23; $p = 0.003$) and the bilateral presentation condition (M = 1.07, SD = 0.26, $p < 0.001$). Furthermore, compatible trials (M = 1.20, SD = 0.39) differed significantly [$p < 0.001$] from incompatible trials (M = 0.74, SD = 0.14) only in the right visual hemifield field condition.[2]

DISCUSSION

The present study aimed at evaluating the postulate of the TCM that the verbal representation of arithmetic facts should be situated unilaterally in the left hemisphere of the human brain (Dehaene and Cohen, 1995, 1997). Therefore, we investigated whether processing of multiplicative triplets in

[2]"Due to the contradictory findings in the literature on the influence of sex on hemispheric asymmetry, in particular for numerical processing (cf. Harris et al., 2018; Pletzer et al., 2019), we have *post hoc* analyzed the data for a potential gender effect. For this analysis, we evaluated influences of lateralized presentation of stimuli on the unit-decade compatibility effect by running a 2 (compatibility: compatible vs. incompatible pairs) × 3 (lateralization: right vs. left vs. bilateral) ANCOVA on the RCS with sex as covariate. The influence of the covariate indicated neither a significant main effect of sex [$F(1,126) = 2.05$, $p = 0.15$, $\eta^2_p = 0.02$] nor an interaction of sex and laterality [$F(2,126) = 0.03$, $p = 0.96$, $\eta^2_p < 0.001$] as well as an interaction of sex and unit-decade compatibility, respectively [$F(,1,126) = 0.22$, $p = 0.64$, $\eta^2_p = 0.002$]. As in the original analysis, there was a marginally significant main effect of compatibility [$F(1,126) = 3.43$, $p = 0.06$, $\eta^2_p = 0.027$], indicating more correctly solved items per second for compatible (M = 1.06, SD = 0.28) as compared to incompatible number pairs (M = 0.93, SD = 0.27). The main effect for lateralization again was not significant [$F(2,126) = 0.39$, $p = 0.67$, $\eta^2_p = 0.0063$]. However, the interaction of compatibility and lateralization was again significant [$F(2,126) = 17.41$, $p < 0.001$, $\eta^2_p = 0.22$].

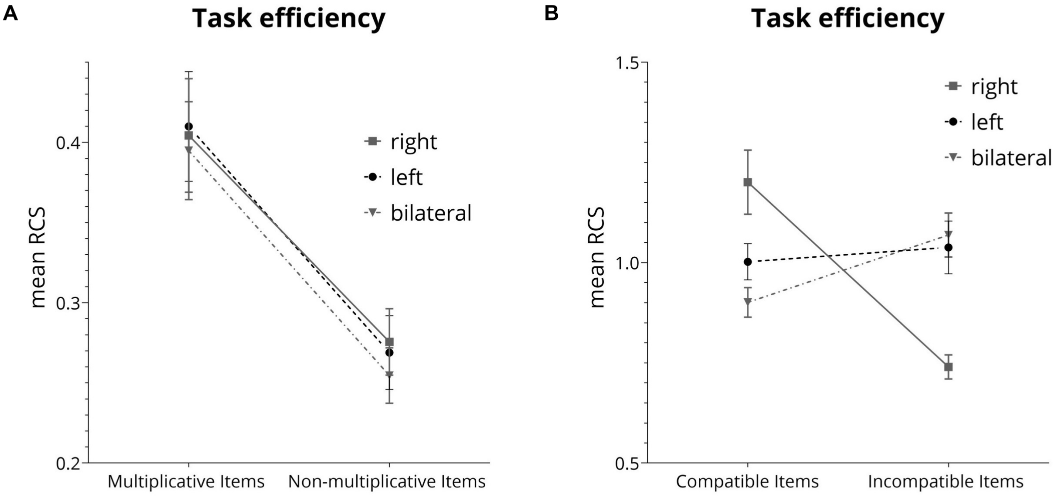


FIGURE 3 | (A) Effect of multiplicativity for the mean rate correct score (RCS). **(B)** Interaction between lateralization conditions and unit-decade compatibility calculated for the mean rate correct score (RCS).

the NBT shows a significant disadvantage when visual input is transmitted to the contralateral right hemisphere only using a divided visual field paradigm. To ensure applicability of the divided visual field paradigm, participants also completed a magnitude comparison task for which influences of lateralized presentation of stimuli was observed previously (Ratinckx et al., 2006).

As regards the latter, we replicated the results by Ratinckx et al. (2006), in particular the modulation of the unit-decade compatibility effect by the lateralization of input presentation in the divided visual field paradigm: presentation of the numbers in the right hemifield (and thus transmitted to the left hemisphere) increased the disadvantage for incompatible number pairs when comparing their magnitude in contrast to the presentation of number pairs in the left visual hemifield or bilaterally.

As regards the NBT, we replicated both standard numerical effects, this means the multiplicativity effect and the bisection possibility effect (e.g., Nuerk et al., 2002; Wood et al., 2008; Moeller et al., 2009, 2011). However, contrary to the hypothesis deriving from the TCM, we did not observe modulation of the multiplicativity effect by lateralization of stimulus presentation. In the following, we will discuss these findings in more detail step by step.

TABLE 2 | Linear regression to check the influence of problem size.

Variable	Estimate	*SE*	*t*-Statistic	*p*-Value
Intercept	8.595e−05	1.767e−05	4.86	<0.001
Multiplicativity[a]	4.941e−04	2.500e−05	19.78	<0.001
Problem size[a]	2.658e−05	2.500e−05	1.06	0.29
Multiplicativity × Problem size	−2.513e−04	3.535e−05	−7.11	<0.001

Adjusted R^2 = 0.66; [a]Factors "Multiplicativity" and "Problem size" have been dummy coded (1 = multiplicative and 1 = large problem size).

Unit-Decade Compatibility Effect in Number Magnitude Comparison

In line with the results of Ratinckx et al. (2006), no particular disadvantage of processing the more difficult incompatible number pairs items was found when stimuli were presented bilaterally or within the left hemifield. Bilateral presentation in both visual fields and unilateral presentation in the left visual hemifield allowed for direct processing of the respective stimuli in the right hemisphere, where the integration of tens and units into the place-value structure of the Arabic number system was argued to take place (Wood et al., 2008). The replication of the results by Ratinckx et al. (2006) suggests that our experimental setting was principally valid for detecting differences caused by the lateralized presentation of stimuli. On a more basic level, these results indicate that participants were able to perceive, and process the presented two-digit numbers even though these were presented only briefly in perifoveal position.

Nevertheless, it has to be noted that participants in our study committed more errors and took longer for their responses as compared to the participants in Ratinckx et al. (2006). While we cannot rule out that this difference might in part be due to unspecific individual differences or cultural differences between participant samples tested (e.g., Dutch vs. German undergraduate students), there were also differences in the experimental setting, which might have contributed to differences in overall accuracy and reaction times of results.

First, Ratinckx et al. (2006) realized lateralization condition block-wise so that within one block all stimuli were presented in the same visual hemifield only. This way, participants knew where to expect the perifoveally presented stimuli (i.e., right, left or bilaterally). By presenting stimuli in different visual hemifields in randomized order, we aimed at preventing attentional orientation toward the left or the right side before the actual stimuli were presented. However, randomized order of stimuli might also have led to longer reaction times and higher errors rates because

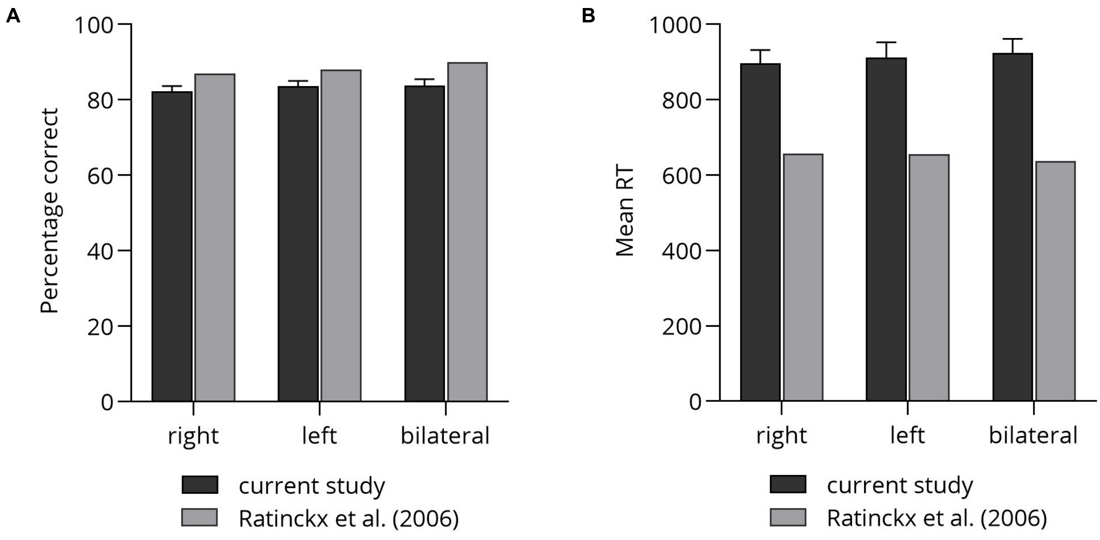


FIGURE 4 | Comparison of participant's test performance: **(A)** mean percentage correct. **(B)** Mean RT. Error bars depict SEM.

participants might have experienced more difficulties in locating and perceiving the stimuli.

Second, we presented visual input in all four locations where a stimulus could potentially be presented (i.e., at each location either one of the two numbers or "##" as a mask was presented). Again, this might have increased task difficulty as more visual input needed to be processed (Poole and Kane, 2009). However, it has to be noted that this was true for all conditions. Thus, it might have potentially affected overall accuracy and reaction time in all conditions to a similar extent, while the differential pattern between lateralized presentation conditions should not have been altered. The latter is reflected by the differential results for the unit-decade compatibility effect depending on lateralization of input presentation.

In sum, we were able to replicate the differential result pattern for the unit-decade compatibility effect as reported by Ratinckx et al. (2006), suggesting that participants were able to perceive and process the two-digit numbers and, more importantly, that our experimental setting was, in principal, valid for detecting differences to due lateralized presentation of numerical stimuli.

Multiplicativity in the Number Bisection Task

Unexpectedly, we did not observe significant modulation of the multiplicativity effect by lateralization of stimulus presentation in the NBT. Generally, this finding allows for two possible conclusions:

First, the assumption of the TCM is wrong, namely that the verbal representation, which underlies arithmetic fact retrieval such as overlearned multiplication facts, is subserved in a lateralized manner in the left hemisphere of the human brain only (cf. Dehaene and Cohen, 1995, 1997; Dehaene et al., 2003). However, before such an interpretation can be considered, other possible explanations need to be ruled out and this finding should be replicated in the same task but also in other tasks drawing on the verbal representation.

Currently, we can only state that multiplicativity as measured in the NBT was not modulated by the site of lateralized presentation in the present experiment. There may be the following reason for this observation: while multiplicativity in the NBT with two-digit number pairs has been argued to draw heavily on verbally mediated arithmetic fact retrieval (cf. Nuerk et al., 2002 for behavioral data; Wood et al., 2008 for neuroimaging data; Klein et al., 2016 for connectivity data), the effect probably may not reflect retrieval of overlearned arithmetic facts only. For instance, whenever participants operate on two-digit numbers, additional processes such as place-value integration (Nuerk et al., 2011) or working-memory (Kong et al., 2005) may be required. In line with this argument, we observed problem size to interact with multiplicativity in the NBT. The interaction specifically indicated that the processing advantage for multiplicative items was smaller for triplets with larger problem size. In turn, this indicates that multiplicativity of triplets was specifically beneficial when triplets consisted of smaller numbers. Such a problem size effect has been reported previously for both children and adults in multiplication (Campbell and Graham, 1985; De Brauwer et al., 2006; but see Domahs et al., 2006). The effect is also in line with the results of Wood et al. (2008) who showed increasing retrieval-specific activation of the left angular gyrus with decreasing problem size of multiplicative triplets in the same version of the NBT as used in the current study. This activation has been repeatedly interpreted to indicate arithmetic fact retrieval (e.g., Dehaene et al., 2003; Delazer et al., 2003; Ischebeck et al., 2006; Grabner et al., 2009). However, as outlined above our experimental setting might have been more difficult than in the study by Wood et al. (2008) due to the brief lateralized perifoveal presentation. When we also take into account the observed differences in overall behavioral performance in magnitude comparison between the study by Ratinckx et al. (2006) and the present study, this possibility can hardly been ruled out. Our way of presenting the NBT might have led to additional demands as compared to previous variants of the NBT in which all numbers were presented simultaneously and in one line. This leads us to the second possible conclusion drawn from our results.

Second, it might be the case that the NBT in its present variant was very and maybe even too difficult for participants. An indicator for this assumption seems the high error rates observed for the NBT as well as the high number of exclusions of participants due to poor performance in the NBT. In particular, from 31 participants, 8 participants had to be excluded for overall scoring below 60% correct in the NBT (with 50% being guessing rate). Additionally, this was combined with a specific pattern of significant better accuracy with slower reaction times in multiplicative triplets. Possibly, the two numbers presented laterally from the fixation cross were perceived and, due to their short presentation duration of only 150 ms, repeated in verbal working memory before a decision on the bisectability of a triplet was made. As multiplicative triplets have been shown to be processed in a verbal code (Moeller et al., 2011), decisions on the multiplicativity of these rehearsed numbers might have been more accurate, as the significant lower error rate for multiplicative triplets may indicate. This would be in line with the idea of better performance in multiplicative triplets in terms of accuracy, while, at the same time, rehearsing three numbers in verbal working memory, would be relatively slow. Support for this assumption also comes from the observation that this specific pattern of higher accuracy synced with slower reaction times was not observed for incorrectly bisected triplets. Thus, the observed behavioral pattern most probably reflects a specific facilitation of the task for multiplicative items as both the multiplication facts as well as verbal working memory operate on a phonological code (Moeller et al., 2011).

While, all these processes are assumed to be subserved by the left hemisphere (Dehaene et al., 2003), we have to consider that we assessed healthy participants with intact interhemispheric connections. Therefore, it will only take a few milliseconds until stimuli may be processed in both hemispheres due to interhemispheric connections via transcallosal fiber pathways (e.g., Chaumillon et al., 2018), so that no temporal processing advantage due to multiplicativity may be observable any more. In other words, the longer processing of the respective stimuli takes, the less likely differences due to lateralized processing in terms of speed should be observed. Therefore, we would suggest that an effect of lateralization of presentation may primarily be expected for early bottom-up effects such as the unit-decade compatibility effect when the task is easy enough. In our magnitude comparison task, only two of the briefly and lateralized presented two-digit numbers were relevant, while participants had to consider three two-digit numbers in the NBT.

In addition, and when interpreting these results, there are two constraints that need to be considered. First, saccadic eye movements were not controlled by eye-tracking; an examiner sitting opposite of the participant monitored eye fixation. Since the center of the lateralized stimuli was located 5 degrees from central fixation and was thus well beyond the critical distance of saccade amplitude that can be detected with the naked eye, we can be very sure that the current procedure has prevented unwanted loss of fixation. Nevertheless, eye-tracking could have been more precise in excluding the possibility that hemispheric asymmetries were only detected when interacting with unit-decade-incompatibility in the number comparison task but not as a main effect of lateralized processing. Second, despite our *a priori* estimation of the necessary sample size to detect hemispheric asymmetry ($N = 21$, a partial eta square of $\eta^2_p = 0.20$ and a power of 0.95), the sample size in the present study ($N = 23$) might have been too small to reveal a main effect of lateralized processing in both the NBT and the number comparison task. However, the observed significant modulation of the compatibility effect by lateralization of stimulus presentation suggests that hemispheric differences are present at least in the magnitude comparison task.

Therefore, it would be desirable for future studies addressing the question of lateralized processing of arithmetic fact retrieval to recruit a larger sample and use easier stimulus material (e.g., one-digit numbers). Moreover, a block-wise realization of lateralized presentation should be applied (e.g., unilateral presentation in the left-hemifield only) in a task which specifically addresses the verbal representation and retrieval of arithmetic facts such as, for instance, one-digit × one-digit multiplications."

Finally, future studies might also evaluate possible influences of lateralized processing in brain areas that cannot be considered independently from lateralized cognitive processing. For instance, the cerebellum has been shown to indirectly regulate activation and inhibition levels of attentional networks (Mannarelli et al., 2019).

CONCLUSION

The current study aimed at investigating whether the verbal representation of arithmetic facts is situated unilaterally in the left hemisphere of the human brain. While we were able to replicate both the multiplicativity effect and the effect of bisection possibility, lateralized presentation did not modulate the effect of multiplicativity.

We suggest that participants might have kept the three two-digit numbers in verbal working memory after perceiving them due to short presentation duration. This would be in line with the observed better performance in multiplicative triplets in terms of accuracy, while, at the same time, reaction times were larger. Rehearsal of the three numbers in the phonological loop was probably too time-consuming to detect fine-grained hemispheric processing asymmetries in multiplicative items due to interhemispheric connectivity. We suggest that an effect of presentation lateralization can only to be expected for early effects such as the unit-decade compatibility effect when the task is easy enough.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee at the Medical Faculty of the Eberhard Karls University and at the University Hospital Tübingen (082/2018BO2). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SJ, KM, H-OK, and EK designed the study. SJ conducted the experiment. SJ and EK analyzed the data and wrote the manuscript. SJ, KM, H-OK, and EK reviewed and approved the final version of the manuscript.

FUNDING

This work was supported the German Research Foundation (KL 2788/2-1 and KA 1258/24-1); and by the Margarete von Wrangell Fellowship (European Social Fund and the Ministry of Science, Research and the Arts Baden-Württemberg).

ACKNOWLEDGMENTS

We wish to thank Roman Janssen for his precious help in data analyzing and all research assistants for their assistance with data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00088/full#supplementary-material

REFERENCES

Bloechle, J., Huber, S., Bahnmueller, J., Rennig, J., Willmes, K., Cavdaroglu, S., et al. (2016). Fact learning in complex arithmetic—the role of the angular gyrus revisited. *Hum. Brain Mapp.* 37, 3061–3079. doi: 10.1002/hbm.23226

Bocci, T., Nasini, F., Caleo, M., Restani, L., Barloscio, D., Ardolino, G., et al. (2018). Unilateral application of cathodal tDCS reduces transcallosal inhibition and improves visual acuity in amblyopic patients. *Front. Behav. Neurosci.* 12:109. doi: 10.3389/fnbeh.2018.00109

Brysbaert, M. (1994). Interhemispheric transfer and the processing of foveally presented stimuli. *Behav. Brain Res.* 64, 151–161. doi: 10.1016/0166-4328(94)90127-9

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

Chaumillon, R., Blouin, J., and Guillaume, A. (2018). Interhemispheric transfer time asymmetry of visual information depends on eye dominance: an electrophysiological study. *Front. Neurosci.* 12:72. doi: 10.3389/fnins.2018.00072

Clemens, B., Zvyagintsev, M., Sack, A. T., Heinecke, A., Willmes, K., and Sturm, W. (2013). Comparison of fMRI activation patterns for test and training procedures of alertness and focused attention. *Restorat. Neurol. Neurosci.* 31, 311–336. doi: 10.3233/RNN-120266

Cohen, L., and Dehaene, S. (2000). Calculating without reading: unsuspected residual abilities in pure alexia. *Cogn. Neuropsychol.* 17, 563–583. doi: 10.1080/02643290050110656

De Brauwer, J., Verguts, T., and Fias, W. (2006). The representation of multiplication facts: developmental changes in the problem size, five, and tie effects. *J. Exp. Child Psychol.* 94, 43–56. doi: 10.1016/j.jecp.2005.11.004

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., Spelke, E., Pinel, P., Stanescu, R., and Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* 284, 970–974. doi: 10.1126/science.284.5416.970

Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Lochy, A., Trieb, T., et al. (2003). Learning complex arithmetic - An fMRI study. *Cogn. Brain Res.* 18, 76–88. doi: 10.1016/j.cogbrainres.2003.09.005

Delazer, M., Ischebeck, A., Domahs, F., Zamarian, L., Koppelstaetter, F., Siedentopf, C. M., et al. (2005). Learning by strategies and learning by drill - Evidence from an fMRI study. *NeuroImage* 25, 838–849. doi: 10.1016/j.neuroimage.2004.12.009

Dimond, S., and Beaumont, G. (1971). Use of two cerebral hemispheres to increase brain capacity. *Nature* 232, 270–271. doi: 10.1038/232270a0

Domahs, F., Delazer, M., and Nuerk, H. C. (2006). What makes multiplication facts difficult: problem size or neighborhood consistency? *Exp. Psychol.* 53, 275–282. doi: 10.1027/1618-3169.53.4.275

Geffen, G., Bradshaw, J. L., and Wallace, G. (1971). Interhemispheric effects on reaction time to verbal and nonverbal visual stimuli. *J. Exp. Psychol.* 87, 415. doi: 10.1037/h0030525

Grabner, R. H., Ischebeck, A., Reishofer, G., Koschutnig, K., Delazer, M., Ebner, F., et al. (2009). Fact learning in complex arithmetic and figural-spatial tasks: the role of the angular gyrus and its relation to mathematical competence. *Hum. Brain Mapp.* 30, 2936–2952. doi: 10.1002/hbm.20720

Granà, A., Hofer, R., and Semenza, C. (2006). Acalculia from a right hemisphere lesion. Dealing with "where" in multiplication procedures. *Neuropsychologia* 44, 2972–2986. doi: 10.1016/j.neuropsychologia.2006.06.027

Harris, T., Scheuringer, A., and Pletzer, B. (2018). Sex differences and functional hemispheric asymmetries during number comparison. *Biol. Sex. Differ.* 9:3. doi: 10.1186/s13293-017-0162-6

Hatta, T., Kawakami, A., Kogure, T., and Itoh, Y. (2002). Effects of type of cognitive demand on bilateral advantage in interhemispheric processing. *Psychol. Res.* 66, 133–142. doi: 10.1007/s004260100077

Hildebrandt, F., Scheidt, A., Hildebrandt, A., Hédervári-Heller, É, and Dreier, A. (2016). *Sustained shared thinking*als Interaktionsformat und das Sprachverhalten von Kindern. *Frühe Bildung* 5, 82–90. doi: 10.1026/2191-9186/a000256

Ischebeck, A., Zamarian, L., Siedentopf, C., Koppelstätter, F., Benke, T., Felber, S., et al. (2006). How specifically do we learn? Imaging the learning of multiplication and subtraction. *NeuroImage* 30, 1365–1375. doi: 10.1016/j.neuroimage.2005.11.016

Klein, E., Moeller, K., Glauche, V., Weiller, C., and Willmes, K. (2013a). Processing Pathways in Mental Arithmetic-Evidence from Probabilistic Fiber Tracking. *PLoS One* 8:e55455. doi: 10.1371/journal.pone.0055455

Klein, E., Moeller, K., and Willmes, K. (2013b). A neural disconnection hypothesis on impaired numerical processing. *Fronti. Hum. Neurosci.* 7:663.

Klein, E., Suchan, J., Moeller, K., Karnath, H. O., Knops, A., Wood, G., et al. (2016). Considering structural connectivity in the triple code model of numerical cognition: differential connectivity for magnitude processing and arithmetic facts. *Brain Struct. Funct.* 221, 979–995. doi: 10.1007/s00429-014-0951-1

Klein, E., Willmes, K., Bieck, S. M., Bloechle, J., and Moeller, K. (2019). White matter neuro-plasticity in mental arithmetic: Changes in hippocampal connectivity following arithmetic drill training. *Cortex* 114, 115–123. doi: 10.1016/j.cortex.2018.05.017

Kong, J., Wang, C., Kwong, K., Vangel, M., Chua, E., and Gollub, R. (2005). The neural substrate of arithmetic operations and procedure complexity. *Cogn. Brain Res.* 22, 397–405. doi: 10.1016/j.cogbrainres.2004.09.011

Mannarelli, D., Pauletti, C., Currá, A., Marinelli, L., Corrado, A., Delle Chiaie, R., et al., (2019). The Cerebellum modulates attention network functioning: evidence from a cerebellar transcranial direct current stimulation and attention network test study. *Cerebellum* 18, 457–468. doi: 10.1007/s12311-019-01014-8

Moeller, K., Fischer, M. H., Nuerk, H. C., and Willmes, K. (2009). Eye fixation behaviour in the number bisection task: evidence for temporal specificity. *Acta Psychol.* 131, 209–220. doi: 10.1016/j.actpsy.2009.05.005

Moeller, K., Klein, E., Fischer, M. H., Nuerk, H. C., and Willmes, K. (2011). Representation of Multiplication Facts-Evidence for partial verbal coding. *Behav. Brain Funct.* 7, 1–9. doi: 10.1186/1744-9081-7-25

Nuerk, H., 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., Geppert, B. E., Van Herten, M., and Willmes, K. (2002). On the impact of different number representations in the number bisection task. *Cortex* 38, 691–715. doi: 10.1016/S0010-9452(08)70038-8

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Pletzer, B., Jäger, S., and Hawelka, S. (2019). Sex hormones and number processing. Progesterone and testosterone relate to hemispheric asymmetries during number comparison. *Horm. Behav.* 115:104553. doi: 10.1016/j.yhbeh.2019.07.001

Poole, B. J., and Kane, M. J. (2009). Working-memory capacity predicts the executive control of visual search among distractors: the influences of sustained and selective attention. *Q. J. Exp. Psychol.* 62, 1430–1454. doi: 10.1080/17470210802479329

Ratinckx, E., Nuerk, H. C., van Dijck, J. P., and Willmes, K. (2006). Effects of interhemispheric communication on two-digit arabic number processing. *Cortex* 42, 1128–1137. doi: 10.1016/S0010-9452(08)70225-9

R Core Team (2019). *R: A Language and Environment For Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: https://www.R-project.org/

Salillas, E., Semenza, C., Basso, D., Vecchi, T., and Siegal, M. (2012). Single pulse TMS induced disruption to right and left parietal cortex on addition and multiplication. *NeuroImage* 59, 3159–3165. doi: 10.1016/j.neuroimage.2011.10.093

Stanescu-Cosson, R., Pinel, P., van de Moortele, P. F., Le Bihan, D., Cohen, L., and Dehaene, S. (2000). Understanding dissociations in dyscalculia: a brain imaging study of the impact of number size on the cerebral networks for exact and approximate calculation. *Brain* 123, 2240–2255. doi: 10.1093/brain/123.11.2240

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: a rejoinder on the binning procedure. *Behav. Res. Methods* 49, 653–673. doi: 10.3758/s13428-016-0721-5

Woltz, D. J., and Was, C. A. (2006). Availability of related long-term memory during and after attention focus in working memory. *Mem. Cogn.* 34, 668–684. doi: 10.3758/BF03193587

Wood, G., Nuerk, H. C., Moeller, K., Geppert, B., Schnitker, R., Weber, J., et al. (2008). All for one but not one for all: how multiple number representations are recruited in one numerical task. *Brain Res.* 1187, 154–166. doi: 10.1016/j.brainres.2007.09.094

Wood, G., Nuerk, H.-C., and Willmes, K. (2006). Neural representations of two-digit numbers: A parametric fMRI study. *NeuroImage* 29, 358–367. doi: 10.1016/j.neuroimage.2005.07.056

Zaunmuller, L., Domahs, F., Dressel, K., Lonnemann, J., Klein, E., Ischebeck, A., et al. (2009). Rehabilitation of arithmetic fact retrieval via extensive practice: a combined fMRI and behavioural case-study. *Neuropsychol. Rehabil.* 19, 422–443. doi: 10.1080/09602010802296378

Conflict of Interest: 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 75 stimuli duplets in the number comparison task.

Within decades	Compatible trials	Incompatible trials
10_16	10_21	17_23
10_14	12_30	16_34
20_29	15_27	22_30
20_24	21_35	24_40
21_27	22_38	25_42
22_26	23_36	27_45
30_39	33_44	34_40
30_36	36_48	38_54
33_39	43_55	39_43
40_45	43_57	39_51
40_49	44_57	44_52
40_46	46_57	45_63
50_56	47_59	48_61
51_57	51_68	52_60
52_58	54_69	58_72
53_57	56_68	59_68
60_65	82_98	64_73
63_69	50_64	68_76
64_69	51_62	72_90
70_76	52_64	74_92
72_78	54_66	76_82
73_79	62_76	77_83
80_88	70_86	87_94
90_99	80_94	63_81
92_98	84_98	69_82

TABLE A2 | Overview of the 180 multiplication problems for the control task.

Correct problems			Incorrect problems		
0 × 2 = 0	4 × 1 = 4	7 × 9 = 63	0 × 2 = 2	4 × 2 = 6	8 × 0 = 8
0 × 3 = 0	4 × 2 = 8	8 × 0 = 0	0 × 3 = 3	4 × 3 = 16	8 × 1 = 7
0 × 4 = 0	4 × 3 = 12	8 × 1 = 8	0 × 4 = 4	4 × 4 = 12	8 × 2 = 8
0 × 7 = 0	4 × 4 = 16	8 × 2 = 16	0 × 7 = 7	4 × 5 = 15	8 × 3 = 21
0 × 8 = 0	4 × 5 = 20	8 × 3 = 24	0 × 8 = 8	4 × 6 = 18	8 × 4 = 28
0 × 9 = 0	4 × 6 = 24	8 × 4 = 32	0 × 9 = 9	4 × 7 = 32	8 × 6 = 42
1 × 2 = 2	4 × 7 = 28	8 × 6 = 48	1 × 2 = 1	4 × 8 = 36	8 × 7 = 48
1 × 4 = 4	4 × 8 = 32	8 × 7 = 56	1 × 4 = 3	5 × 1 = 6	8 × 8 = 56
1 × 5 = 5	5 × 1 = 5	8 × 8 = 64	1 × 5 = 4	5 × 2 = 15	8 × 9 = 64
1 × 7 = 7	5 × 2 = 10	8 × 9 = 72	1 × 7 = 6	5 × 3 = 20	9 × 0 = 9
1 × 8 = 8	5 × 3 = 15	9 × 0 = 0	1 × 8 = 9	5 × 4 = 24	9 × 1 = 18
1 × 9 = 9	5 × 4 = 20	9 × 1 = 9	1 × 9 = 0	5 × 5 = 20	9 × 2 = 27
2 × 0 = 0	5 × 5 = 25	9 × 2 = 18	2 × 0 = 2	5 × 6 = 36	9 × 3 = 24
2 × 1 = 2	5 × 6 = 30	9 × 3 = 27	2 × 1 = 3	5 × 7 = 30	2 × 4 = 12
2 × 2 = 4	5 × 7 = 35	4 × 10 = 40	2 × 2 = 2	5 × 9 = 40	4 × 10 = 30
2 × 3 = 6	5 × 9 = 45	5 × 10 = 50	2 × 3 = 4	6 × 9 = 45	5 × 10 = 45
2 × 4 = 8	6 × 9 = 54	6 × 10 = 60	2 × 5 = 8	7 × 0 = 7	6 × 10 = 50
2 × 5 = 10	7 × 0 = 0	7 × 10 = 70	2 × 6 = 6	7 × 1 = 8	7 × 10 = 77
2 × 6 = 12	7 × 1 = 7	8 × 10 = 80	3 × 4 = 9	7 × 3 = 24	8 × 10 = 72
3 × 4 = 12	7 × 3 = 21	10 × 3 = 30	3 × 5 = 12	7 × 4 = 21	10 × 3 = 20
3 × 5 = 15	7 × 4 = 28	10 × 4 = 40	3 × 7 = 28	7 × 5 = 42	10 × 4 = 50
3 × 7 = 21	7 × 5 = 35	10 × 5 = 50	3 × 8 = 32	7 × 6 = 48	10 × 5 = 41
3 × 8 = 24	7 × 6 = 42	10 × 6 = 60	3 × 9 = 36	7 × 7 = 42	10 × 6 = 54
3 × 9 = 27	7 × 7 = 49	10 × 7 = 70	4 × 0 = 4	7 × 8 = 64	10 × 7 = 60
4 × 0 = 0	7 × 8 = 56	10 × 8 = 80	4 × 1 = 5	7 × 9 = 72	10 × 8 = 90

frontiers in Psychology

ORIGINAL RESEARCH
published: 26 March 2020
doi: 10.3389/fpsyg.2020.00549

Check for updates

Symbolic Processing Mediates the Relation Between Non-symbolic Processing and Later Arithmetic Performance

Sabrina Finke[1]*, H. Harald Freudenthaler[1] and Karin Landerl[1,2]

[1] Institute of Psychology, University of Graz, Graz, Austria, [2] Department of Cognitive Science, Macquarie University, Sydney, NSW, Australia

The nature of the relation between non-symbolic and symbolic magnitude processing in the prediction of arithmetic remains a hotly debated subject. This longitudinal study examined whether the influence of non-symbolic magnitude processing on arithmetic is mediated by symbolic processing skills. A sample of 130 children with age-adequate ($N = 73$) or below-average ($N = 57$) achievement in early arithmetic was followed from the end of Grade 1 (mean age: 86.9 months) through the beginning of Grade 4. Symbolic comparison of one- and two-digit numbers serially mediated the effect of non-symbolic comparison on later arithmetic. These results support a developmental model in which non-symbolic processing provides a scaffold for single-digit processing, which in turn influences multi-digit processing and arithmetic. In conclusion, both non-symbolic and symbolic processing play an important role in the development of arithmetic during primary school and might be valuable long-term indicators for the early identification of children at risk for low achievement in arithmetic.

Keywords: numerical cognition, non-symbolic, symbolic, longitudinal, mediation

OPEN ACCESS

Edited by:
Reuven Babai,
Tel Aviv University, Israel

Reviewed by:
Louise A. Brown Nicholls,
University of Strathclyde,
United Kingdom
Xinlin Zhou,
Beijing Normal University, China

****Correspondence:***
Sabrina Finke
sabrina.finke@uni-graz.at

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *11 October 2019*
Accepted: *09 March 2020*
Published: *26 March 2020*

Citation:
Finke S, Freudenthaler HH and Landerl K (2020) Symbolic Processing Mediates the Relation Between Non-symbolic Processing and Later Arithmetic Performance. Front. Psychol. 11:549. doi: 10.3389/fpsyg.2020.00549

INTRODUCTION

The development of arithmetic skills in primary school is of fundamental importance in modern-day societies: already at the age of seven, arithmetic abilities predict adult socio-economic status over and above the effects of intelligence and socio-economic status at birth (Ritchie and Bates, 2013). Severe deficits in arithmetic are relatively stable: almost half of the children diagnosed with developmental dyscalculia at an age of 11 still meet the diagnostic criteria 6 years later (Shalev et al., 2005). Therefore, it is important to discover the cognitive mechanisms underlying arithmetic achievement in order to identify and support children at risk before their problems get persistent. However, longitudinal studies unraveling the effects of different, interacting predictors of the development of arithmetic are still scarce (Alcock et al., 2016).

Children's arithmetic development has often been linked to their "number sense," meaning the ability to deal with non-symbolic magnitudes, for example dots or other concrete objects. Typical tasks involve choosing the numerically larger of two sets of objects (e.g., ●● or ●●●). This ability has been proposed to reflect the acuity of the supposedly innate approximate number system (ANS). Using a habituation-dishabituation methodology, it became apparent that 6-month-olds can differentiate between sets with a ratio of 1:2 (Xu and Spelke, 2000). During child development, non-symbolic skills are steadily refined, until young adults can successfully discriminate between sets with a ratio of 10:11 (Halberda and Feigenson, 2008).

It has been argued that non-symbolic magnitude processing is directly and causally related to arithmetic performance (Dehaene, 2002; Halberda et al., 2008). Support for this claim is mostly derived from correlational studies showing that non-symbolic processing skills are related to past, concurrent or future arithmetic performance (Halberda et al., 2008; Gilmore et al., 2010; Libertus et al., 2011). Additionally, it has been proposed that developmental dyscalculia is the result of an inborn "core deficit" of acuity of non-symbolic processing (Wilson and Dehaene, 2010).

Others have rejected the notion of a causal relation between non-symbolic magnitude processing and arithmetic, alternatively proposing that the ability to deal with abstract symbolic numbers (mainly in the form of digits) is more important for arithmetic performance. Symbolic magnitude processing is often assessed with tasks requiring participants to indicate which of two Arabic digits is numerically larger (e.g., 2 or 3). In this vein, compared to typical development, children with dyscalculia showed lower performance in a symbolic magnitude comparison task, but not when comparing non-symbolic numerosities (Rousselle and Noël, 2007). Based on a systematic review, De Smedt et al. (2013) concluded that symbolic processing is a more robust predictor of arithmetic than non-symbolic processing, as many studies failed to find a significant correlation between non-symbolic magnitude comparison and arithmetic. This assumption was recently confirmed by two meta-analyses (Fazio et al., 2014; Schneider et al., 2017) reporting a significantly stronger association with mathematics for symbolic than for non-symbolic magnitude processing. In studies that assess both, symbolic and non-symbolic processing, the latter typically does not contribute additional variance to the prediction of arithmetic over and above symbolic processing (Lyons and Ansari, 2015) leading some researchers to the conclusion that non-symbolic processing skills are "not particularly critical for children's development of school-relevant mathematical competencies" (De Smedt et al., 2013, p. 54).

The fact that non-symbolic processing skills do not explain additional variance in arithmetic performance when controlling for differences in symbolic processing does not dismiss a potentially causal relation between non-symbolic processing and arithmetic. Only recently, an alternative mediation hypothesis has proposed that the relation between non-symbolic processing skills and arithmetic might be mediated by symbolic skills (Lyons and Beilock, 2011; Fazio et al., 2014; van Marle et al., 2014; Price and Fuchs, 2016; Peng et al., 2017; Träff et al., 2018). An evolutionary based ability to discriminate between sets of objects may provide a starting point for young children's mapping of numerical symbols (number words, Arabic numbers) onto non-symbolic numerosities, which in turn are the foundation of their arithmetic skills (Dehaene, 2002). Nevertheless, empirical support for the claim that non-symbolic skills provide a scaffold for symbolic skills, which in turn predict arithmetic performance, is mostly based on cross-sectional studies.

For instance, Lyons and Beilock (2011) found that ordering skills fully mediated the association between non-symbolic processing skills and arithmetic in a sample of young adults. The authors argued that the ability to comprehend the relative order of digits might be grounded in an ANS and act as a stepping stone for the acquisition of arithmetic skills in a small sample (N = 53) of fifth graders, Fazio et al. (2014) found that composite scores of non-symbolic and symbolic processing independently contributed to the prediction of mathematics. When the authors tested for a possible indirect effect by examining the reduction of the direct effect of non-symbolic skills on arithmetic performance once symbolic processing was added. This reduction of the direct effect just about missed significance. Thus, there is at least some evidence for a weak indirect effect which might have well been significant if the sample size had been larger.

If non-symbolic processing is a foundation of understanding symbolic numbers, it might be expected that it is of particular relevance in young children who are still developing their symbolic number system. Indeed, Peng et al. (2017) reported for a sample of kindergarten children aged five to six that a composite "numerical knowledge" variable significantly mediated the relation between non-symbolic processing skills and arithmetic, even when controlling for a variety of covariates, including intelligence, working memory, attention and inhibition. Numerical knowledge consisted of rapid automatized naming with digits, identification of one- to three-digit numbers, and numerical reasoning (completing a sequence of numbers). Similarly, Price and Fuchs (2016) found full concurrent mediation of the relation between non-symbolic processing and arithmetic by symbolic processing in a sample of 9-year-olds, even when working memory skills were controlled for. In order to keep task requirements as similar as possible, non-symbolic and symbolic skills were both assessed by comparison tasks encompassing numerosities from 1 to 9. Importantly, non-symbolic processing did not conversely mediate the effect of symbolic processing on arithmetic. In a similar age group of third graders, Träff et al. (2018) also found evidence for an indirect effect of non-symbolic processing speed on single-digit arithmetic mediated by symbolic processing speed, over and above the influence of linguistic skills, as indexed by language comprehension and rapid automatized naming. Non-symbolic processing skills were measured with a computerized task comprising numerosities between 5 and 21 per array (Panamath; Halberda et al., 2008), whereas the symbolic processing measure consisted of a composite score of single- and double-digit comparisons. In another study with 5- to 8-year-olds (Li et al., 2018) non-symbolic as well as symbolic processing tasks were assessed in numerosities from 5 to 50. Interestingly, the effect of non-symbolic processing on mathematical ability was mediated by symbolic processing skills in children aged 5–6, but not in 7- to 8-year-olds, providing first evidence that the age of assessment may be critical.

Longitudinal studies are particularly relevant in order to determine causal mechanisms during development. So far, only one such study (van Marle et al., 2014) investigated symbolic skills as a potential mediator of the relation between non-symbolic magnitude processing in 3- to 4-year-olds on entering preschool and mathematic abilities at the end of the preschool year. The mathematical abilities test encompassed items involving enumeration, counting, cardinal knowledge and numeral identification, but critically, calculation skills or arithmetic fact knowledge could not be assessed in this young age group, which may well explain why symbolic magnitude comparison was not found to be a mediator for the relation between non-symbolic processing and these very basic mathematical competences.

In summary, several studies investigating the hypothesis that symbolic processing abilities serve as a mediator of the relation between non-symbolic processing skills and arithmetic did indeed report some evidence in favor of this claim. Conflicting findings might in part be due to the different measures of symbolic number processing that were employed (e.g., ordinality judgment, numerical recognition, and number comparison) and differences in age groups assessed. Still, the foundational link between non-symbolic and symbolic processing is not entirely uncontested: while there is some evidence that non-symbolic processing influences the development of symbolic processing skills (Toll et al., 2015), other studies could not corroborate this link and reported that, on the contrary, symbolic skills predicted growth in non-symbolic processing (Mussolin et al., 2014; Matejko and Ansari, 2016; Lyons et al., 2018). Therefore, it appears crucial that studies testing a mediation model of non-symbolic processing on arithmetic via symbolic processing should control for initial symbolic skills, in order to test whether non-symbolic processing actually predicts growth in symbolic processing skills.

In addition, it might also be important to differentiate between distinct subcomponents of symbolic processing, in particular single- and multi-digit number processing. There is increasing evidence suggesting that multi-digit number processing differs from single-digit processing and is acquired later in development (Brankaer et al., 2017). Moreover, it has been proposed that single-digit number processing constitutes a necessary first step, while additional specific processes, such as place-value knowledge, are required to fully understand multi-digit numbers (Nuerk et al., 2015). Thus, it seems plausible to assume that the ability to process multi-digit numbers is scaffolded onto single-digit number processing, which in turn may rely on non-symbolic processing.

In the current study we tested a developmental model of sequential mediation of the effect of non-symbolic processing on later arithmetic performance via processing of one- and two-digit numbers. This developmental account was investigated in a longitudinal study ranging from end of Grade 1 to beginning of Grade 4. In this important period of arithmetic development, children are introduced to the complexities of the Arabic place-value system and are expected to acquire fluent competencies in mental calculation and to store a large amount of easily accessible number facts in their long-term memory.

Non-symbolic, single- and multi-digit number processing were assessed at different, sequential time points, and prior to arithmetic skills. When testing our developmental framework, we controlled for general cognitive skills that have been found to be associated with arithmetic performance and might influence its relation with non-symbolic processing, i.e., non-verbal IQ (Göbel et al., 2014b), verbal working memory (Berg, 2008), and attention/executive functions (Clark et al., 2013). Furthermore, as the children had already gained substantial experience with symbolic numbers and arithmetic at the beginning of our study period, it was important to additionally control for initial symbolic and arithmetic skills.

Based on the findings of previous studies, we expect that non-symbolic processing exerts an indirect effect of future arithmetic performance but may not uniquely contribute to the prediction of arithmetic performance (i.e., no total effect) when considering these control variables. Note that this study design puts the hypothesis that non-symbolic processing is a foundational skill underlying symbolic processing and arithmetic at a very stringent test: while non-symbolic processing could be expected to have its strongest influence early in development, when numbers are mapped onto analog magnitudes, our longitudinal design mainly assesses whether differences in non-symbolic processing contribute to growth of symbolic processing and (in turn) arithmetic skills during the primary school years, over and above general cognitive predictors of arithmetic. If any such indirect long-term effects can be demonstrated, even though small, the hypothesis that the non-symbolic magnitude processing is a foundational skill of arithmetic should be further investigated.

MATERIALS AND METHODS

Participants

The study was conducted in accordance with the ethical principles of the World Medical Association Declaration of Helsinki. Data collection started in 2007 and at that time the funding agency (DFG) and local legislation did not request an explicit vote from an ethics committee for non-medical research. Legal guardians gave their written informed consent before data collection. The present sample consisted of 130 children from 19 different elementary schools and a total of 38 classrooms taking part in a longitudinal study investigating the developmental trajectories of basic numerical skills in children with typical and atypical arithmetic development (Landerl, 2013).

The participants were invited to the study based on a screening of 505 children at the end of first grade. Children with arithmetic achievement of 1 *SD* or more below age norm on a standardized test (Haffner et al., 2005) were all invited for additional assessments. For each participant with below-average arithmetic achievement, we selected one child from the same classroom who displayed typical arithmetic development (i.e., arithmetic performance above −1 *SD* compared to the age norm). Thus, children with low arithmetic performance in Grade 1 were overselected in our sample. As our focus was on numerical and arithmetic development, we attempted to exclude more general deficits in non-verbal IQ, working memory, attention, and reading as potential causes or confounds of arithmetic deficits. More specifically, children were not admitted to the study if they met any of the following exclusion criteria:

- a native language other than German;
- IQ lower than 85 as assessed by a test of non-verbal intelligence (Cattell et al., 1997);
- verbal working memory more than 1 *SD* below age norm on the German version of the WISC-IV digit span subtest (Petermann and Petermann, 2008);
- a clinical diagnosis of attention deficit/hyperactivity disorder or performance more than 1.5 *SD*s below age norm on a standardized test of attention/executive functions (Zimmermann et al., 2002);
- reading abilities more than 2 *SD*s below age norm, as measured by a standardized reading test (Mayringer and Wimmer, 2003). As the deficits of children with co-occurring dyslexia and dyscalculia appear to be additive but not qualitatively different from isolated disorders (Landerl et al., 2009), a conservative cut-off for reading problems was chosen.

The initial sample consisted of 139 children (68 boys and 71 girls), of whom 131 participated through Grade 4. One child with low arithmetic performance had to be excluded because of below chance-level performance on the non-symbolic comparison task (rendering it unclear whether this child had understood the instruction). Thus, the final sample comprised 130 children (60 boys and 70 girls) with an average age of 86.9 months at the screening (end of Grade 1). At this first assessment point, 57 children showed low arithmetic performance. At the last assessment point in Grade 4, only 39 children performed more than 1 *SD* below the age norm in arithmetic, while the majority of the sample ($N = 91$) showed arithmetic skills within the typical range.

Design

Children's development was followed across a 2-year primary school period from end of Grade 1 (T1) to beginning of Grade 4 (T4). The first assessment point (T1) subsumed measures that were either given at the end of Grade 1, or right at the beginning of Grade 2, interspersed only by 6 weeks school holidays. Non-symbolic processing at T1 was considered as independent predictor variable. Symbolic single-digit processing was assessed 6 months later in the middle of Grade 2 (T2) and symbolic two-digit processing was assessed after another 6 months at the beginning of Grade 3 (T3). The dependent variable arithmetic performance was assessed at the beginning of Grade 4 (T4). We additionally considered several covariates: non-verbal intelligence, attention, verbal working memory, as well as initial arithmetic performance and symbolic single-digit processing (all T1, except for attention, which was assessed at T2 because of restricted assessment time at T1). An overview of the study design is depicted in **Figure 1**.

Tasks

Numerical Processing

Non-symbolic and symbolic processing were assessed by standard numerical comparison paradigms programmed with Presentation software. Children performed the tasks individually in a quiet room at their school. We obtained a combined measure of speed and accuracy for children's non-symbolic and symbolic processing skills by calculating inverse efficiency scores (median reaction times divided by the proportion of correct responses).

Non-symbolic comparison

Children were required to indicate which of two gray displays had the larger number of yellow squares (see **Figure 2**) by pressing the corresponding keyboard button as rapidly as possible. The number of squares per display ranged from 20 to 72 squares in order to discourage children from explicit counting. The difference between displays' set sizes ranged from 8 to 25 squares, with four trials for each numerical distance, resulting in a total of 72 test items. The total surface area was the same on both displays, and the same proportion of both displays was covered by yellow squares. Each display consisted of different square sizes to avoid that displays with larger numerosities systematically consisted of smaller squares. The largest and smallest squares appeared in the same number in both displays; only size and number of intermediate squares were different. Stimuli were displayed in a fixed pseudo-random order and remained on screen until the child made a keypress decision. After an interstimulus-interval of 300 ms, the next item appeared. At the start, three practice items were presented. Cronbach's alpha reliability for the non-symbolic comparison task was 0.95.

Symbolic comparison

Two tasks assessed symbolic processing skills: single digit comparison (T1 and T2) and comparison of two-digit numbers (T3). In both tasks, children chose the numerically larger of two numbers by pressing the corresponding keyboard button as quickly as possible. The single digit comparison task consisted of 56 items with numbers from 1 to 9. Numerical distances ranged from 1 to 8 (distance 1: 16 items, distances 2–3: 10 items, and distances 4–8: 4 items). Comparison of two-digit numbers comprised 80 items with numbers between 21 and 98. Numerical distance ranged from 4 to 37. In 30 items, both decade and unit digit were larger in one number (e.g., 41 75), in 30 items, the decade digit was larger in one and the unit digit was larger in the other number (e.g., 41 26), and in further 20 items, only the unit digit differed (e.g., 61 68). In both symbolic processing tasks, stimuli were displayed in a 36-point Times New Roman font in black color against a white background. Item presentation was randomized and the number pairs remained on the screen until children responded by keypress. After each item, there was an interstimulus-interval of 560 ms. For both symbolic comparison measures, Cronbach's alpha ranged between 0.96 and 0.97.

Arithmetic Performance

Arithmetic performance was assessed by the "arithmetic operations" subscale of a standardized classroom test (Haffner et al., 2005). At T1, the assessment included lists of addition, subtraction, fill-in-the-blank (e.g., $10 - 2 = 4 + \_$) and size comparison exercises (e.g., $51 - 1 \_ 6$; fill in " > "). At T4, two additional subtests targeting multiplication and division were included. Each subtest had a 2 min time limit with items being presented with increasing difficulty. Within this time limit, children were required to write down as many correct answers as possible to a list of calculations gradually increasing

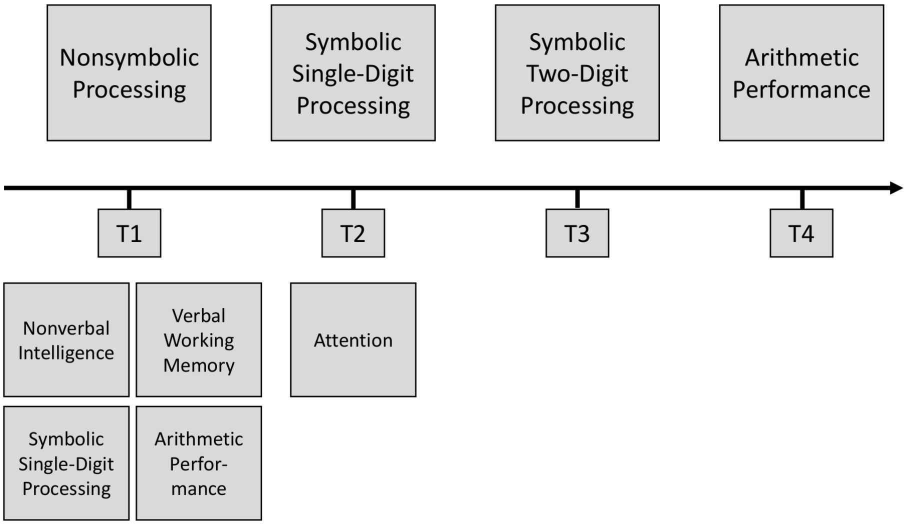


FIGURE 1 | Overview of the study design. Main study variables are depicted above the timeline and covariates are depicted below the timeline.

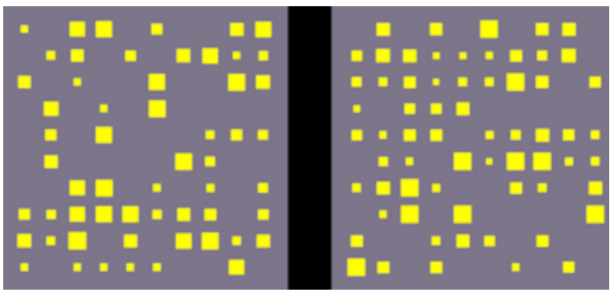


FIGURE 2 | Example item of the magnitude comparison task (Landerl, 2013).

in difficulty. Performance in each subtest was assessed as the number of correct answers. A composite measure "arithmetic operations" was calculated as the mean of the standardized *T*-scores (mean = 50, *SD* = 10) of all subtests.

Non-verbal Intelligence

Non-verbal intelligence was assessed by the German version of the Culture Fair Test (Cattell et al., 1997), comprising the subtests substitutions, mazes, classifications, similarities and matrices. These five subtests provided a measure of general intellectual ability, i.e., a child's ability to recognize regularities and quickly identify characteristics.

Attention

Children performed a standardized computer-based test battery encompassing different facets of attention/executive functions (Zimmermann et al., 2002). Attention was indexed by a composite score of the subtests distractibility, alertness, sustained attention, flexibility, and divided attention. In the distractibility subtest, children were required to selectively press a button upon seeing a ghost with a sad face. In half of the trials, a distractor in form of a ghost with a happy face appeared right before the target stimulus. In the alertness subtest, a witch appeared in the center of the screen at varying intervals and children had to press a button as quickly as possible. The sustained attention subtest measured children's ability to maintain their attention over a longer period of time (10 min) by watching the color of ghosts that appeared on the screen one after the other. They had to press a button whenever two subsequent ghosts had the same color. In the flexibility subtest, a green and a blue dragon appeared simultaneously on the screen, and children had to indicate the positions of both dragons. In alternating trials, the position of the green versus the blue dragon had to be indicated first. During the divided attention subtest, children were presented with different visual and auditory stimuli. They had to react to changes in the stimuli, i.e., when an owl closed its eyes or changed its hooting.

Verbal Working Memory

The Digit Span subtest (forward and backward combined) of the German version of the Wechsler Intelligence Scale for Children IV (Petermann and Petermann, 2008) was used to assess verbal working memory. In the forward condition, children were required to repeat a string of verbal numbers presented by the experimenter in the same order. In the backward condition, they were asked to repeat the number strings in the inverse order. For each number length, two items were presented and a discontinuation rule applied if a child was unable to repeat at least one of these items. Verbal working memory was indexed as the total score of correctly recalled number strings.

RESULTS

For each of the numerical processing tasks, individual median response times were calculated after removing reaction times for incorrect responses, below 200 ms and above 10,000 ms.

In the non-symbolic comparison task the correlation between median RTs and response accuracy was only moderate, $r = 0.324$, $p < 0.001$ and response accuracy was close to ceiling in both symbolic comparison tasks (mean accuracy symbolic comparison with single-digit numbers at T1 = 95.5% and T2 = 96.6% and symbolic comparison with two-digit numbers at T3 = 91.3%). In order to combine response accuracy and speed, inverse efficiency scores were computed for the non-symbolic and symbolic tasks by dividing the median reaction time by the proportion of correct responses. Finally, one extreme outlier score in the non-symbolic comparison task (more than 6 *SD*s above the sample mean) was moved to the tail of the distribution to the second highest score to avoid overemphasizing its effect on the results.

Descriptive Statistics

Means, standard deviations and intercorrelations of all relevant study variables are shown in **Table 1**. Pearson correlation coefficients were reported to describe the linear relations between study variables. If a pair of those variables was not bivariately normally distributed, confidence intervals obtained by bootstrapping and Spearman coefficients were also computed to examine statistical significance. In all of these cases, the three approaches yielded identical results regarding statistical significance.

Mediation Analyses

Mediation analyses were calculated using the PROCESS macro for SPSS (Hayes, 2013). In order to evaluate the hypothesis that the influence of non-symbolic magnitude processing on later arithmetic performance is sequentially mediated by symbolic magnitude processing of single- and multi-digit numbers, we calculated a serial multiple mediation analysis with bootstrapping. During the bootstrapping procedure, the current sample was randomly resampled with replacement. An empirically obtained representation of the sampling distribution of the indirect effect was used to generate the confidence interval for the indirect effect. In the current study, we employed a bias-corrected bootstrap with a 95% confidence intervals based on 10,000 bootstrap samples.

Residualized change scores of arithmetic (arithmetic T4 – arithmetic T1) were considered as dependent variable, and symbolic processing of single- and two-digit numbers were introduced as mediators, so we obtained the following PROCESS model: non-symbolic processing T1 → symbolic processing (single-digit numbers) T2 → symbolic processing (two-digit numbers) T3 → residualized change in arithmetic T4-T1. We added several general cognitive covariates of both mediators and the dependent variable, namely verbal working memory, non-verbal intelligence and attention. We also included a fourth covariate, symbolic processing of single-digit numbers at T1. The full PROCESS model including the main standardized path coefficients is depicted in **Figure 3**. The only significant direct effects were from non-symbolic to symbolic single-digit processing (*a1* path) and from single-digit to two-digit processing (*a3* path). Importantly, neither non-symbolic processing skills at T1 nor symbolic single-digit processing at T2 exerted significant direct effects on arithmetic growth (c' and *b1* paths). The direct paths from non-symbolic to two-digit symbolic processing (*a2*) and from two-digit symbolic processing to arithmetic growth (*b2*) missed significance (*p*s = 0.061 and 0.054, respectively).

Indirect effects, their standard errors and confidence intervals are presented in **Table 2**. Non-symbolic processing did not exert a significant indirect effect on arithmetic growth through symbolic processing of single-digit numbers at T2 (*a1b1* path) or through symbolic processing of two-digit numbers at T3 (*a2b2* path). Still, non-symbolic processing did show a significant (though small) influence on arithmetic growth serially through symbolic processing of single-digit numbers at T2 and symbolic processing of two-digit numbers at T3 (*a1a3b2* path).

DISCUSSION

Non-symbolic Magnitude Processing and Arithmetic

The presented sequential mediation analyses indicated that the effect of non-symbolic processing in Grade 2 on arithmetic performance 2 years later, in Grade 4, was sequentially mediated by symbolic magnitude processing of one- and two-digit numbers. Even though this mediation effect was small, these results provide an important empirical contribution to the ongoing debate whether non-symbolic processing skills make a causal contribution to arithmetic development. Our developmental perspective on the association of magnitude processing with arithmetic (see also Verguts and Fias, 2004; von Aster and Shalev, 2007) is consistent with evidence that symbolic processing is more strongly associated with arithmetic than non-symbolic processing (De Smedt et al., 2013; Schneider et al., 2017). The fact that the contribution of non-symbolic processing to later arithmetic is indirect via symbolic processing skills can explain why non-symbolic processing did not account for variance above and beyond symbolic processing in earlier studies (Göbel et al., 2014b; Lyons and Ansari, 2015). Our results support the theoretical view that early non-symbolic processing skills make a small but significant contribution to later arithmetic performance. Importantly however, this contribution is completely indirect by providing a scaffold for symbolic processing of one- and two-digit numbers.

Regarding the hypothesis that non-symbolic processing influences arithmetic via symbolic processing, the current longitudinal evidence provides support for the causal claims made based on cross-sectional studies (Price and Fuchs, 2016; Peng et al., 2017; Träff et al., 2018). By covering a relatively long period of over two critical years of early mathematical development and considering a variety of possible confounding factors, we extended the findings by van Marle et al. (2014) on their kindergarten sample. We found a significant direct contribution to symbolic processing of single-digit numbers and an almost significant direct contribution to symbolic processing of double-digit numbers. Importantly, we found a significant long-term indirect contribution of non-symbolic processing skills to arithmetic growth toward the end of the primary school period. A further distinctive feature of our design was that we controlled for early differences in non-verbal IQ, verbal working

TABLE 1 | Descriptive statistics and correlations between all relevant study variables.

Variable	Mean	*SD*	1	2	3	4	5	6	7	8	9
(1) Arithmetic T1[a]	45.50	9.11	–								
(2) Non-symbolic T1[b]	1686.16	353.41	−0.29**	–							
(3) Non-verbal Intelligence T1[c]	108.25	11.71	0.49**	−0.22*	–						
(4) Symbolic Single-Digit T1[b]	1103.76	247.56	−0.41**	0.47**	−0.24**	–					
(5) Verbal Working Memory T1[b]	11.80	1.82	0.31**	−0.06	0.18*	−0.06	–				
(6) Attention T2[a]	47.03	4.46	0.46**	−0.31**	0.39**	−0.45**	0.20*	–			
(7) Symbolic Single-Digit T2[b]	985.04	216.75	−0.42**	0.47**	−0.36**	0.69**	0.01	−0.38**	–		
(8) Symbolic Two-Digit T3[b]	45.50	9.11	−0.43**	0.46**	−0.35**	0.62**	−0.08	−0.33**	0.62**	–	
(9) Arithmetic T4[a]	1686.16	353.41	0.72**	−0.27**	0.51**	−0.40**	0.22*	0.41**	−0.48**	−0.50**	–

*[a]Mean T-Score of all Subtests (M:50/SD:10).[b]Inverse Efficiency Score. [c]IQ Score (M:100/SD:15). *p < 0.05; **p < 0.01.*

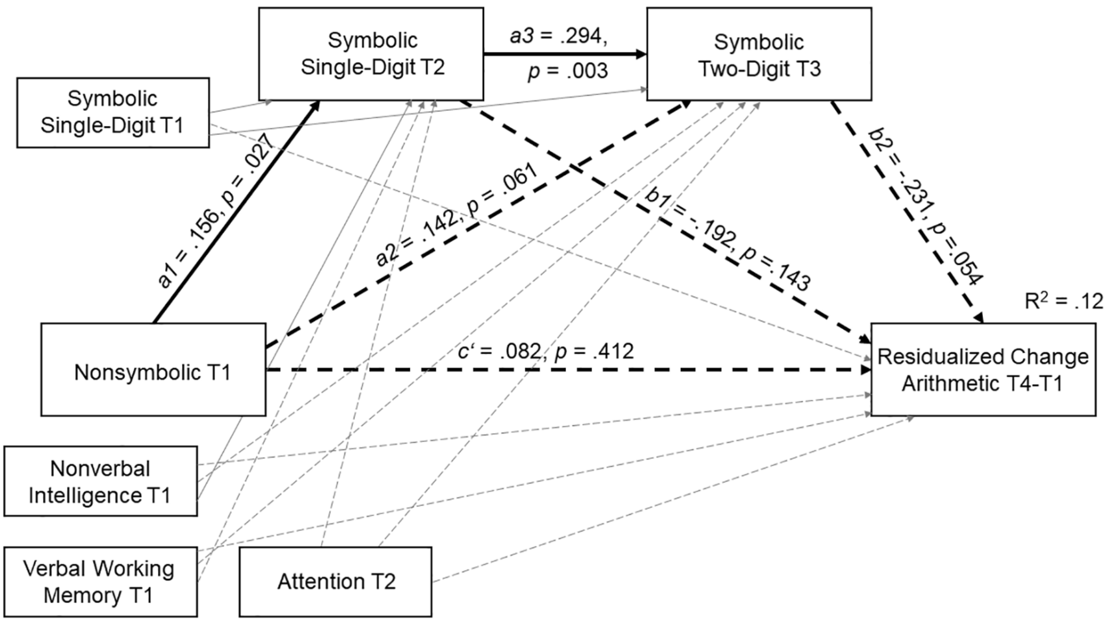


FIGURE 3 | Serial multiple mediation model for the effect of non-symbolic processing on residualized change scores in arithmetic performance T4-T1 with symbolic processing of single- and two-digit numbers as mediators.

TABLE 2 | Effects, standard errors, and bootstrapped confidence intervals of non-symbolic processing on residualized change scores in arithmetic between T1 and T4 (controlling for non-verbal intelligence, attention, verbal working memory, and symbolic magnitude processing at T1, contributing to the mediators and arithmetic performance).

Effects			Estimate (*SE*)	LCI	UCI
Direct:	*c′*	Non-symbolic T1 → Arithmetic Growth T4 – T1	0.082 (0.100)	−0.112	0.280
Indirect:	*a1b1*	Non-symbolic T1 → Symbolic Single-Digit T2 → Arithmetic Growth T4 – T1	−0.030 (0.027)	−0.113	0.003
	a1a3b2	Non-symbolic T1 → Symbolic Single-Digit T2 → Symbolic Two-Digit T3 → Arithmetic Growth T4 – T1	−0.011 (0.011)	−0.057	−0.001
	a2b2	Non-symbolic T1 → Symbolic Two-Digit T3 → Arithmetic Growth T4 – T1	−0.033 (0.032)	−0.135	0.005
Total:		Non-symbolic T1 Arithmetic Growth T4 – T1	0.008 (0.099)	−0.118	0.205

LCI, lower CI bound; UCI, upper CI bound.

memory and attention. As expected and consistent with earlier research (Berg, 2008; Clark et al., 2013; Göbel et al., 2014b), these general cognitive factors were significantly related to growth in arithmetic skills across the study period. It is particularly impressive that the mediation pathway from early non-symbolic magnitude processing to growth in arithmetic was significant across these critical years of primary school and beyond the influence of these general cognitive predictors and even after controlling for interindividual differences in single-digit symbolic processing at the onset of the study period. Given the design of our study, it is not particularly surprising that this effect was numerically small and one could argue that it is irrelevant as its ecological validity is low. However, from a theoretical point of view, such a small but significant long-term effect suggests that current proposals that non-symbolic processing may be entirely irrelevant for understanding symbolic representations of number and arithmetic (De Smedt et al., 2013) are perhaps premature. On the contrary, our findings encourage further research on the exact mechanisms underlying the associations of non-symbolic processing with symbolic numerical processing and arithmetic skills from a developmental perspective.

Non-symbolic and Symbolic Processing

In contrast to a number of recent studies, we found a significant contribution of early non-symbolic processing to growth in symbolic processing skills half a year later even after controlling for a variety of general cognitive variables. A number of other longitudinal studies with kindergarten and first grade children failed to find a similar contribution (Mussolin et al., 2014; Matejko and Ansari, 2016; Lyons et al., 2018). It is not unlikely that the special characteristics of the sample investigated here increased the chance to reveal such a relation. At the onset of the study, almost half of our participants were selected because they showed early problems in arithmetic performance. Therefore, the variance in non-symbolic and symbolic processing skills was perhaps larger than in randomly selected samples, which may have helped to reveal a relation that is small and therefore hard to detect in the normal population. It is also possible that this association is only evident in individuals with deficits in arithmetic development. This would be in line with assumptions that there may be two subtypes of dyscalculia: one with a core deficit in non-symbolic magnitude processing and another one with intact magnitude processing but problems to access magnitude representations from symbolic number representations (Rousselle and Noël, 2007). Depending of the profiles of individual participants within a sample, findings may vary. Unfortunately, our sample was too small to run separate analyses for children with arithmetic deficits and as a matter of fact, only a subgroup of those with early problems turned out to develop persistent deficits in arithmetic. In future studies, it might be worthwhile to investigate whether the early relation between non-symbolic processing and growth in symbolic processing as well as the observed indirect effect of sequential mediation between non-symbolic processing and later arithmetic may be specific to dyscalculia.

Arabic Number Processing and Arithmetic

The sequential mediation model presented here also critically extends empirical evidence on the pivotal role of understanding single- and multi-digit Arabic numbers for arithmetic development. As predicted, the ability to process multi-digit numbers was scaffolded onto single-digit number processing. This finding supports the proposed developmental trajectory.

As pinpointed previously (Nuerk et al., 2015), being able to deal with single-digit numbers is an important prerequisite, but perhaps not sufficient for multi-digit number processing. Understanding the relation between decade and unit position is one of the additional steps required for two-digit number processing. It is interesting that in the current study the direct contribution of non-symbolic processing to two-digit number processing missed significance ($p = 0.061$). Future research should test the hypothesis that the ability to represent non-symbolic magnitudes facilitates the acquisition of place-value understanding.

The finding that in our model single-digit processing at T2 (middle of Grade 2) did not show a direct influence on arithmetic growth from T1 to T4 is probably due to the fact that we controlled for differences in symbolic processing at the onset of the study period. This means that the single-digit processing variable actually only reflects changes in task performance through a period of about 6 months. Variance in this variable was predicted by non-symbolic processing skills half a year earlier and in turn contributed significantly to processing of two-digit numbers. Its contribution to arithmetic growth from Grade 2 to Grade 4 was, however, indirect as a mediator of non-symbolic processing.

Although a strong association between performance in single- and two-digit comparison tasks was found in the present study (see also Brankaer et al., 2017), these tasks appear to measure distinct constructs and contribute differently to the prediction of arithmetic performance. Future studies on the development of arithmetic should therefore ideally include both single- and multi-digit number processing tasks. Only few studies have so far dealt with the development of multi-digit number processing (Nuerk et al., 2015). Given the increasing evidence on the high relevance of understanding place-value and multi-digit syntax for arithmetic development (Moeller et al., 2009; Moura et al., 2013; Göbel et al., 2014a), it will be important to investigate the particular challenges children are facing when acquiring complex Arabic numbers and their verbal counterparts. These challenges are mathematical (place-value) as well as linguistic (e.g., inversion of 10s and units in German and other languages) and seem to constitute an important milestone in arithmetic development.

Limitations

There has been an ongoing discussion on how to best assess non-symbolic magnitude processing skills (Price et al., 2012; Schneider et al., 2017). In the present design, we prioritized having similar tasks for non-symbolic and symbolic processing in order to rule out any confounding effects of differences in task format. Other measures that have been claimed to be more sensitive (e.g., Weber fraction), might have produced stronger effects in our mediation model than the combined measure of accuracy and speed introduced here (including a potential direct effect on later arithmetic).

Similarly, it has been claimed that tasks specifically tapping into cardinal or ordinal number knowledge might be better mediators between non-symbolic processing and mathematical skills. We consider it highly plausible that development of counting plays a crucial role in the mapping of number words and Arabic digits onto non-symbolic processing skills (van Marle et al., 2014). However, although we covered a relatively long developmental period in our longitudinal design, our data do not address the very early foundational processes of this mapping process.

We would also like to remind readers that the current sample was not randomly selected: at the onset of the study period, half of the participants were selected based on below-average performance in arithmetic development. It turned out that later on the majority of children displayed typical arithmetic performance. Still, the distribution of numerical processing and arithmetic skills may not correspond to the general population and replication with more representative samples is advisable.

CONCLUSION

In summary, the evidence presented in our study reveals a significant role of non-symbolic magnitude processing in the development of arithmetic during primary school: we could demonstrate that non-symbolic processing skills impacted on growth of arithmetic skills by facilitating the acquisition of symbolic number processing. This evidence indicates that non-symbolic processing should be included as one of the foundational skills in theoretical models of mathematical development. It will also be important to further specify developmental trajectories within the domain of symbolic numerical processing, by, for instance, differentiating between simple processing of one-digit numbers and more complex processes involved in multi-digit processing.

As the indirect effect exerted by non-symbolic processing was small, it seems unlikely that interventions exclusively targeting non-symbolic processing would show satisfactory effects on children's arithmetic development (Szücs and Myers, 2017). However, drill-practicing number knowledge without providing children with sufficient opportunities to understand how numbers represent non-symbolic magnitudes may be equally inefficient. Training programs should thus focus on understanding and efficiently processing symbolic representation of number, which entails to map them on their inborn non-symbolic representational system (e.g., Kuhn and Holling, 2014; Rauscher et al., 2016). Improving our knowledge of longitudinal developmental trajectories and neurocognitive mechanisms of symbolic and arithmetic processing skills is necessary in order to further advance our understanding of the components that should be integrated in evidence-based tailored intervention programs.

DATA AVAILABILITY STATEMENT

The dataset analyzed for this study can be found on the Open Sciene Framework at https://osf.io/dfbqj/.

ETHICS STATEMENT

The study was conducted in accordance with the ethical principles of the World Medical Association Declaration of Helsinki. Legal guardians of the participating children gave their written informed consent before data collection.

AUTHOR CONTRIBUTIONS

KL developed the study concept and study design and is responsible for data collection. SF performed the data analysis and interpretation under the supervision of HF. SF drafted the manuscript. HF and KL provided critical revisions. All authors approved the final version of the manuscript for submission.

REFERENCES

Alcock, L., Ansari, D., Batchelor, S., Bisson, M.-J., De Smedt, B., Gilmore, C., et al. (2016). Challenges in mathematical cognition: a collaboratively-derived research agenda. *J. Numer. Cogn.* 2, 20–41. doi: 10.5964/jnc.v2i1.10

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

Brankaer, C., Ghesquière, P., and De Smedt, B. (2017). Symbolic magnitude processing in elementary school children: a group administered paper-and-pencil measure (SYMP Test). *Behav. Res. Methods* 49, 1361–1373. doi: 10.3758/s13428-016-0792-3

Cattell, R. B., Weiß, R. H., and Osterland, J. (1997). *Grundintelligenztest Skala 1 (CFT 1) [Basic Intelligence Scale 1]*. Göttingen: Hogrefe.

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

De Smedt, B., Noël, M. P., Gilmore, C., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills 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

Dehaene, S. (2002). Single-neuron arithmetic. *Science* 297, 1652–1653. doi: 10.1126/science.1076392

Fazio, L. K., Bailey, D. H., Thompson, C. A., and Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *J. Exp. Child Psychol.* 123, 53–72. doi: 10.1016/j.jecp.2014.01.013

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

Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H. C. (2014a). Language affects symbolic arithmetic in children: the case of number word inversion. *J. Exp. Child Psychol.* 119, 17–25. doi: 10.1016/j.jecp.2013.10.001

Göbel, S. M., Watson, S. E., Lervag, A., and Hulme, C. (2014b). Children's arithmetic development: it is number knowledge, not the approximate number sense, that counts. *Psychol. Sci.* 25, 789–798. doi: 10.1177/0956797613516471

Haffner, J., Baro, K., Parzer, P., and Resch, F. (2005). *Heidelberger Rechentest (HRT 1-4) [Heidelberg Calculation Test]*. 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., 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

Hayes, A. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, NY: Guilford Press.

Kuhn, J.-T., and Holling, H. (2014). Number sense or working memory? The effect of two computer based trainings on mathematical skills in elementary school. *Adv. Cogn. Psychol.* 10, 59–67. doi: 10.5709/acp-0157-2

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

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

Li, Y., Zhang, M., Chen, Y., Deng, Z., Zhu, X., and Yan, S. (2018). Children's non-symbolic and symbolic numerical representations and their associations with mathematical ability. *Front. Psychol.* 9:1035. doi: 10.3389/fpsyg.2018.01035

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

Lyons, I. M., and Ansari, D. (2015). Foundations of children's numerical and mathematical skills: the roles of symbolic and nonsymbolic representations of numerical magnitude. *Adv. Child Dev. Behav.* 48, 93–116. doi: 10.1016/bs.acdb.2014.11.003

Lyons, I. M., and Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition* 121, 256–261. doi: 10.1016/j.cognition.2011.07.009

Lyons, I. M., Bugden, S., Zheng, S., De Jesus, S., and Ansari, D. (2018). Symbolic number skills predict growth in nonsymbolic number skills in kindergarteners. *Dev. Psychol.* 54, 440–457. doi: 10.1037/dev0000445

Matejko, A. A., and Ansari, D. (2016). Trajectories of symbolic and nonsymbolic magnitude processing in the first year of formal schooling. *PLoS One* 11:e0149863. doi: 10.1371/journal.pone.0149863

Mayringer, H., and Wimmer, H. (2003). *Salzburger Lesescreening (SLS 1-4) [Salzburg Reading Screening]*. Bern: Hans Huber.

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

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

Mussolin, C., Nys, J., Content, A., and Leybaert, J. (2014). Symbolic number abilities predict later approximate number system acuity in preschool children. *PLoS One* 9:e91839. doi: 10.1371/journal.pone.0091839

Nuerk, H.-C., Moeller, K., and Willmes, K. (2015). "Multi-digit number processing: overview, conceptual clarifications, and language influences," in *The Oxford Handbook of Numerical Cognition*, eds R. C. Kadosh and A. Dowker (Oxford: Oxford Library of Psychology), 106–139. doi: 10.1093/oxfordhb/9780199642342.013.021

Peng, P., Yang, X., and Meng, X. (2017). The relation between approximate number system and early arithmetic: the mediation role of numerical knowledge. *J. Exp. Child Psychol.* 157, 111–124. doi: 10.1016/j.jecp.2016.12.011

Petermann, F., and Petermann, U. (2008). *Hamburg-Wechsler-Intelligenztest für Kinder IV (HAWIK-IV) [Wechsler Intelligence Scale for Children IV]*. Bern: Huber.

Price, G. R., and Fuchs, L. S. (2016). The mediating relation between symbolic and nonsymbolic foundations of math competence. *PLoS One* 11:e0148981. doi: 10.1371/journal.pone.0148981

Price, G. R., Palmer, D., Battista, C., and Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychol.* 140, 50–57. doi: 10.1016/j.actpsy.2012.02.008

Rauscher, L., Kohn, J., Käser, T., Mayer, V., Kucian, K., McCaskey, U., et al. (2016). Evaluation of a computer-based training program for enhancing arithmetic skills and spatial number representation in primary school children. *Front. Psychol.* 7:913. doi: 10.3389/fpsyg.2016.00913

Ritchie, S. J., and Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychol. Sci.* 24, 1301–1308. doi: 10.1177/0956797612466268

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

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372

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. doi: 10.1017/S0012162205000216

Szücs, D., and Myers, T. (2017). A critical analysis of design, facts, bias and inference in the approximate number system training literature: a systematic review. *Trends Neurosci. Educ.* 6, 187–203. doi: 10.1016/j.tine.2016.11.002

Toll, S. W. M., Van Viersen, S., Kroesbergen, E. H., and Van Luit, J. E. H. (2015). The development of (non-) symbolic comparison skills throughout kindergarten and their relations with basic mathematical skills. *Learn. Individ. Differ.* 38, 10–17. doi: 10.1016/j.lindif.2014.12.006

Träff, U., Olsson, L., Skagerlund, K., and Östergren, R. (2018). Cognitive mechanisms underlying third graders' arithmetic skills: expanding the pathways to mathematics model. *J. Exp. Child Psychol.* 167, 369–387. doi: 10.1016/j.jecp.2017.11.010

van Marle, K., Chu, F. W., Li, Y., and Geary, D. C. (2014). Acuity of the approximate number system and preschoolers' quantitative development. *Dev. Sci.* 17, 492–505. doi: 10.1111/desc.12143

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

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. (2010). "Number sense and developmental dyscalculia," in *Human Behavior, Learning, and the Developing Brain*, eds D. Coch, G. Dawson, and K. W. Fischer (New York, NY: Guilford Press), 212–238.

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). *Testbatterie zur Aufmerksamkeitsprüfung für Kinder (KITAP) [Test Battery for Attentional Performance in Children]*. Herzogenrath: Psytest.

Conflict of Interest: 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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 15 April 2020
doi: 10.3389/fnhum.2020.00137

Neural Processing Mechanism of Mental Calculation Based on Cerebral Oscillatory Changes: A Comparison Between Abacus Experts and Novices

Abdelkader Nasreddine Belkacem[1,2]*, *Kanako Kiso*[3], *Etsuko Uokawa*[3], *Tetsu Goto*[3], *Shiro Yorifuji*[3] and *Masayuki Hirata*[2,3,4,5]*

[1]Department of Computer and Network Engineering, College of Information Technology, United Arab Emirates University, Al Ain, United Arab Emirates, [2]Department of Neurological Diagnosis and Restoration, Osaka University Graduate School of Medicine, Suita, Japan, [3]Department of Neurosurgery, Osaka University Graduate School of Medicine, Suita, Japan, [4]Endowed Research Department of Clinical Neuroengineering, Global Center for Medical Engineering and Informatics, Osaka University, Suita, Japan, [5]Center for Information and Neural Networks (CiNet), National Institute of Information and Communications Technology, Osaka University, Suita, Japan

OPEN ACCESS

Edited by:
*Korbinian Moeller,
University of Tübingen, Germany*

Reviewed by:
*José Luis Ulloa,
University of Talca, Chile
Patrick Müller,
German Center for Neurodegenerative Diseases (DZNE), Germany*

***Correspondence:**
*Abdelkader Nasreddine Belkacem
belkacem@uaeu.ac.ae
Masayuki Hirata
mhirata@ndr.med.osaka-u.ac.jp*

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *06 November 2019*
Accepted: *23 March 2020*
Published: *15 April 2020*

Citation:
Belkacem AN, Kiso K, Uokawa E, Goto T, Yorifuji S and Hirata M (2020) Neural Processing Mechanism of Mental Calculation Based on Cerebral Oscillatory Changes: A Comparison Between Abacus Experts and Novices. Front. Hum. Neurosci. 14:137. doi: 10.3389/fnhum.2020.00137

Background: Abacus experts could mentally calculate fast some mathematical operations using multi-digit numbers. The temporal dynamics of abacus mental calculation are still unknown although some behavioral and neuroimaging studies have suggested a visuospatial and visuomotor neural process during abacus mental calculation. Therefore, this contribution aims to clarify the significant similarities and the differences between experts and novices by investigating calculation-induced neuromagnetic responses based on cerebral oscillatory changes.

Methods: Twelve to 13 healthy abacus experts and 17 non-experts participated in two experimental paradigms using non-invasive neuromagnetic measurements. In experiments 1 and 2, the spatial distribution of oscillatory changes presented mental calculations and temporal frequency profiles during addition while examining multiplication tasks. The MEG data were analyzed using synthetic aperture magnetometry (SAM) with an adaptive beamformer to calculate the group average of the spatial distribution of oscillatory changes and their temporal frequency profiles in source-level analyses.

Results: Using a group average of the spatial distribution of oscillatory changes, we observed some common brain activities in both right-handed abacus experts and non-experts. In non-experts, we detected the right dorsolateral prefrontal cortex (DLPFC) and bilateral Intraparietal sulcus (IPS); whereas in experts, detected the bilateral parieto-occipital sulcus (POS), right inferior frontal gyrus (IFG), and left sensorimotor areas mainly. Based on the findings generated, we could propose calculation processing models for both abacus experts and non- experts conveniently.

Abbreviations: MEG, Magnetoencephalography; AMC, Abacus-based mental calculation; IPS, Intraparietal sulcus; POS, Parieto-occipital sulcus; IFG, Inferior frontal gyrus; DLPFC, Dorsolateral prefrontal cortex; mPFC, The medial prefrontal cortex; ERS, Event-related synchronization; ERD, Event-related desynchronization; SAM, Synthetic aperture magnetometry.

Conclusion: The proposed model of calculation processing in abacus experts and novices revealed that the novices could calculate logically depending on numerical processing in the left IPS. In contrast, abacus experts are utilizing spatial processing using a memorized imaginary abacus, which distributed over the bilateral hemispheres in the IFG and sensorimotor areas.

Keywords: abacus mental calculation, cerebral oscillatory, brain hemispheres activation, neural mechanism, magnetoencephalography, synthetic aperture magnetometry

INTRODUCTION

For decades, the abacus-based mental calculation has been a unique Asian culture for a long time for rapid and precise calculations. For instance, old Asian people have relied on physical devices, such as the abacus or *Soroban* in Japanese, to perform complex computations. The experts in the abacus can perform some complex computations mentally with fast speed of response and high accuracy of the answer. However, these neural bases of computation processing are still not precisely known; especially, the neural processing mechanism based on cerebral oscillations (e.g., oscillatory changes in the frequency bands, as alpha (8–13 Hz), beta (13–30 Hz) and gamma (>30 Hz) bands). These oscillatory changes associated with specific functional roles (i.e., cognitive processes) over given brain areas (Dimitriadis et al., 2016). The designation of "Abacus experts" refers to those who have gained an unusual ability to operate an abacus for mathematical operations, as well as calculating mentally with an abacus in mind after almost daily training throughout the years.

Training on the abacus-based mental calculation (AMC) has received much attention in neuroscience communities for some clinical and non-clinical applications (Tanaka et al., 2002; Hanakawa et al., 2003; Chen et al., 2006; Hu et al., 2011; Li et al., 2013; Wang et al., 2013, 2017). Most researchers have been trying to understand how the brain works when someone uses an abacus to gain arithmetic skills (Dehaene et al., 1990, 1999, 2003; Dehaene, 1992, 1996, 1997, 2008; Dehaene and Cohen, 1997). There is psychological evidence that abacus experts utilize a mental image of an abacus to recall and manipulate large numbers in solving calculation problems; however, the neural correlates underlying this expertise is still unknown (Cohen et al., 1997, 2004; Cohen Kadosh et al., 2007, 2008), and the mathematical language has always been compared to natural language (Amalric and Dehaene, 2016, 2019). Usually, if someone asks abacus experts "how they pull through any mental calculation, they all say, "We do the calculation by using an abacus within my brain."

Former scholarly studies have covered some relevant active brain areas using positron-emission tomography (PET) and functional magnetic resonance imaging (fMRI) for detecting brain activities related to mathematical and calculation processing, while the importance of the parietal lobe and frontal lobe had always been visible (Dehaene, 1996; Cohen et al., 2000; Rickard et al., 2000). However, since the information obtained from PET and fMRI is based mainly on changes in blood flow, metabolism, and secondary to electrophysiological activities, the time resolution is not high, and it is difficult to capture the reaction at the earliest time-based latency. There are also some prior experimental studies (e.g., Pauli et al., 1994; Iguchi and Hashimoto, 2000) used electroencephalogram (EEG) to measure brain electrical activity directly. In contrast, it is still difficult to capture the mental calculation-based reaction time by using the EEG technique due to its low spatial resolution and high-frequency components.

In this research study, we employed magnetoencephalography (MEG) to measure specific brain activity during mental calculation, where its signals considered as cerebral rhythm changes. The aperture synthetic magnetometry (SAM) method using a nonlinear beamforming approach analyzed the generated MEG signals to investigate brain activity during calculation processing (Taniguchi et al., 2000; Robinson et al., 2004). The SAM method is characterized by canceling noise using a spatial filter and having a high spatial resolution. Furthermore, it is possible to obtain the brain activity region associated with each task by setting the region of interest in the cranium to a lattice shape (voxel) and statistically comparing the signal intensities before and after performing the task in each voxel. Some studies based on MEG have been focusing on decoding the processing stage of mental arithmetic calculations at the single-trial level (Pinheiro-Chagas et al., 2018).

We focused the attention on abacus experts who have sound skills to numbers, along with general brain computation processing mechanisms of healthy subjects. For instance, In Japan, the abacus has been staying in use as a calculation device for a long time. When training an abacus for many years, calculation speed would be faster than the untrained person; so, calculation results could be derived with a high correct answer rate. Also, several reports indicate that the ability to manipulate numbers in memory is excellent, and some prior studies were focusing on special computing abilities (e.g., high calculation processing speed) to have concluded that abacus experts are visual-spatial learners (e.g., Tanaka et al., 2002; Hanakawa et al., 2002; Chen et al., 2006; Wu et al., 2009). To the best of our knowledge; however, mental calculation based on oscillatory changes has not been studied thoroughly using magnetoencephalography.

Although it is being clarified that calculations are made to, it has not been thoroughly studied yet using the magnetoencephalography to the best of our knowledge. Therefore, in this study, we have been conducting research aiming to clarify the difference between the computational processing mechanism of the skilled abacus and non-skilled

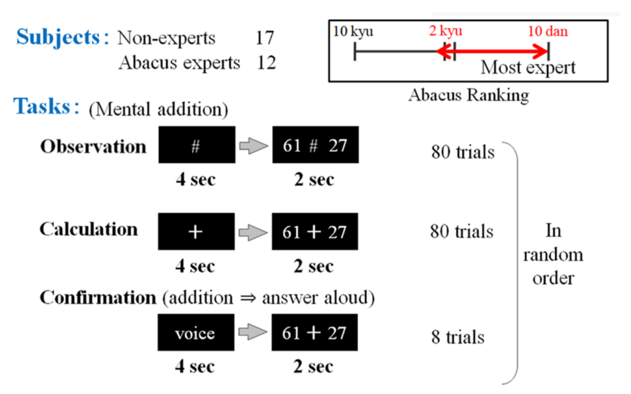


FIGURE 1 | The first experimental paradigm. The participants (17 non-experts and 12 abacus experts) were instructed to perform or confirm at least a mental addition of 2 digits plus 2 digits' number, immediately after numbers were presented (e.g., 61 + 27). Participants (novices and experts in using a Soroban) were asked to perform three tasks: observation, calculation, and confirmation.

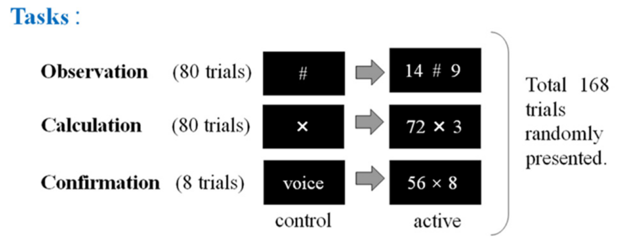


FIGURE 2 | The second experimental paradigm. The participants (17 non-experts and 13 abacus experts) were instructed to perform, observe, or confirm some arithmetic operations by multiplying a two-digit number by a one-digit number (e.g., 72 × 3), immediately after numbers were presented. Participants (novices and experts in using a Soroban) were asked to perform three tasks: calculation, observation, and confirmation. "X" means the execution of mental calculation and "#" means just observation of numbers.

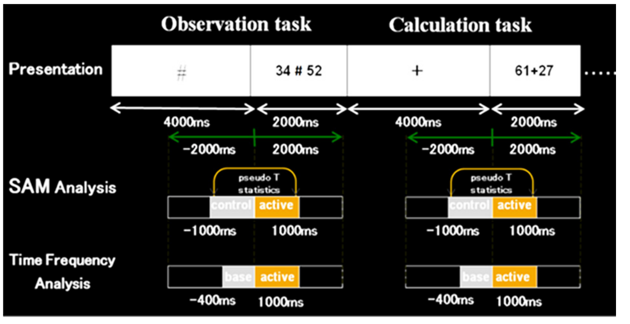


FIGURE 3 | Schematic diagram of the first experimental paradigm. This figure shows a schematic diagram of the task presentation, and a time window of synthetic aperture magnetometry and time-frequency analyses. First, in control stimuli, we presented "+" or VOICE for 4 s, which means what the subject should do in active stimuli. Then, in active stimuli, numbers for calculation were presented for 2 s. In SAM analysis, we compared 1 s before and after the presentation of numbers. Also, regarding time-frequency analysis, the baseline was set for 400 ms before the number presentation.

persons using magnetoencephalography. So far, research has been done on the difference in the calculation process when mentalizing the addition, and non-experts sequentially calculate mental arithmetic with numeric morphological recognition, numerical processing, numeric inner words, working memory, calculation execution processing. While processing, abacus experts got the result that they were doing through several processes, including internal language and calculation processing, at the same time.

However, the additional task used in this study was very easy for experts, and the load on the brain might be substantially different for each subject. Therefore, we used multiplication tasks too to select additional and multiplication tasks based on the level of difficulty, which could be calculated within a particular time by carrying out preliminary experiments and used them in this study which makes the burden on the brain for each subject became equal. The primary driver of this research study is not only clarifying the localization analysis by using the SAM method but also the time-frequency analysis and clarifying the processing process in the brain during multiplication and mental arithmetic.

Therefore, we investigated calculation-induced neuromagnetic responses based on cerebral oscillatory changes. These oscillatory changes are now widely used for functional neuroimaging studies. The present study aims to clarify the spatiotemporal distribution of the cerebral oscillatory changes during mental calculations using synthetic aperture magnetometry (Ihara et al., 2003a,b; Hirata et al., 2007, 2010) and to elucidate the processing mechanism of mental calculation to elucidate the difference between abacus experts and non-experts using magnetic source imaging from magnetoencephalography (MEG) signals. Understanding the neural mechanism of abacus might lead to enhance the calculation ability for patients with Acalculia who are unable to perform mathematical calculations, although some alternative numerical processes are still available to them. This article presents the first model based on the temporal frequency profile of oscillatory changes. Two groups of abacus experts and non-experts were asked to perform some mental calculations to analyze the similarities and differences between them by looking to their temporal profiles to design neural processing models for abacus and non-abacus experts then we compared our results to previous studies (Cohen et al., 1997; Dehaene et al., 1999, 2003; Ishii et al., 1999).

MATERIALS AND METHODS

Participants

Healthy volunteers (17 non-experts from 21 to 55 years of age, and 12–13 abacus experts from 18 to 35) participated in this study to investigate calculation-induced neuromagnetic responses based on cerebral oscillatory changes using non-invasive measurement. For mental addition (+) experiments, the age of non-expert participants is from 21 to 55 years old (Average ± Standard deviation: 24.9 ± 7.86) and the age of experts is from 19 to 24 years old (21.9 ± 2.84). Only

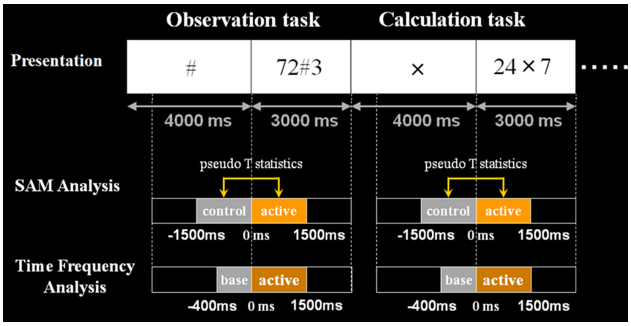


FIGURE 4 | Schematic diagram of the second experimental paradigm. This figure shows the schematic diagram of the task presentation and analyses the time window. First, in control stimuli, we presented X, # or VOICE for 4 s, which means what the subject should do in active stimuli. Then, in active stimuli, numbers for calculation were presented for 3 s. In SAM analysis, we compared 1.5 s before and after the presentation of numbers. Also, regarding time-frequency analysis, the baseline was set for 400 ms before the number presentation. Analyzed frequency bands are as shown here.

one non-expert volunteer was middle-aged (55 years old). For multiplication (×) experiments, the age of non-expert participants is from 21 to 55 years old (25.7 ± 9.3) and the age of experts is from 18 to 35 years old (23.5 ± 3.7). However, there was no statistically significant age difference between participants in both mental calculation experiments, which means that there is no age effects. All participants (experts and non-experts) are right-handed. Abacus experts were certified by the authority for their skills, which were ranging from 2 Kyu to 10 Dan according to the Japanese abacus ranking system (i.e., One Dan is one's degree or level of expertise and knowledge). These experts got abacus training for 6–27 years in their life, and their abacus ranking varies from one person to another.

All participants who were informed in detail about the research purpose and possible consequences of the MEG experiment have signed upon an explicit written consent. The Ethics Committee at Osaka University Hospital approved the conduct of this study, and the experimental protocol was carried out according to the latest version of the Declaration of Helsinki. The T1 structural MRI scans were performed to obtain DICOM images of the head and brain structures in slices for all participants. The acquisition of individual anatomical MRIs of participants was combined with MEG data for getting more precise source localization.

Experimental Paradigm and Protocol

We performed the MEG recording and MRI at the Osaka University Hospital (Japan). Neuromagnetic brain activity was measured with a 64-channel MEG system equipped

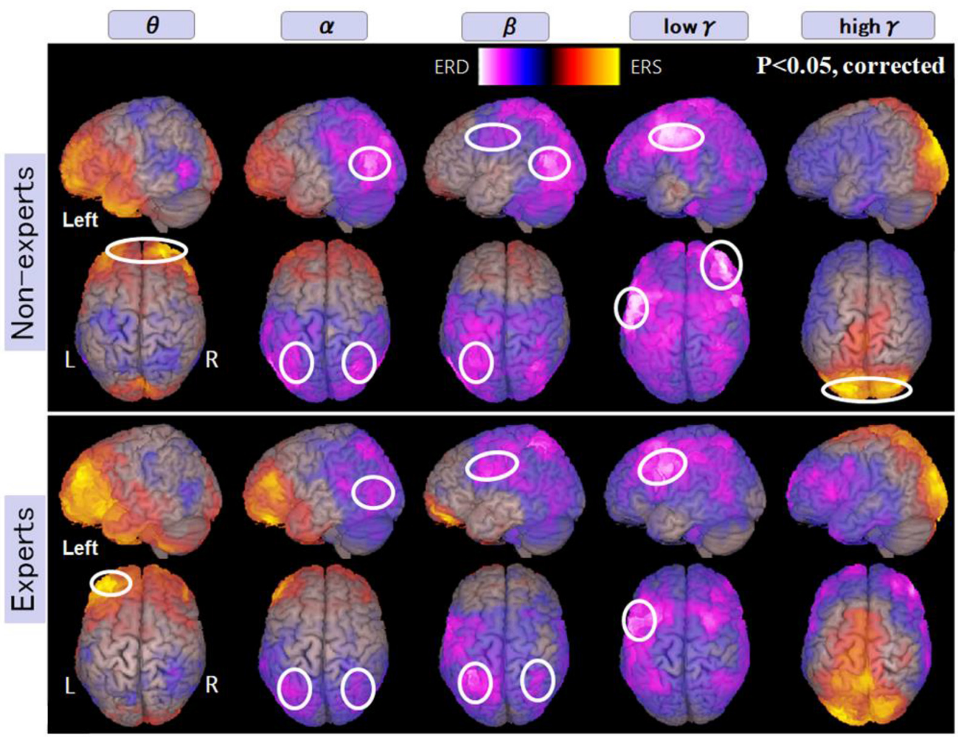


FIGURE 5 | A group average of the spatial distribution of oscillatory changes for the first experiment (mental addition during calculation task). The time window is between −1,000 and 1,000 ms. The frequency interval (all frequency bands) ranges from 4 Hz to 100 Hz. Magenta color shows ERD and orange color shows ERS. The circled areas indicate statistically-significant oscillatory changes. The significant differences observed ($p < 0.05$, corrected) in some brain areas are surrounded by a white circle.

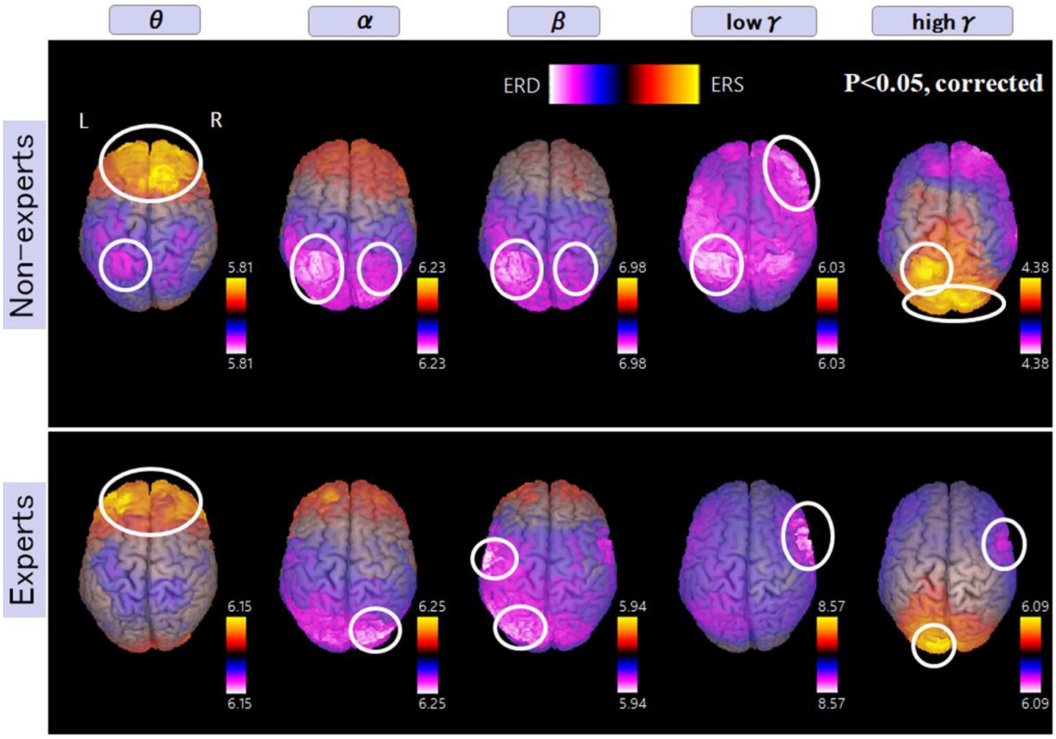


FIGURE 6 | A group average of the spatial distribution of oscillatory changes for the second experiment (mental multiplication during calculation task). The time window is between −1,500 and 1,500 ms. The frequency interval (all frequency bands) ranges from 4 Hz to 100 Hz. Magenta color shows ERD and orange color shows ERS. The circled areas indicate statistically-significant oscillatory changes. The significant differences observed ($p < 0.05$, corrected) in some brain areas are surrounded by a white circle.

with the whole-head array of first-order radial SQUID gradiometers (NeuroSQUID Model 100, CTF Systems Inc., Canada).

We performed two experiments under simple and complex conditions based on the three mental-operation tasks; these were: (i) numeral or calculation; (ii) observation; and (iii) verbal or confirmation (Hanakawa et al., 2002). The participant was in a sitting position, and a projection screen was fixed in front of the eyes. Visual stimuli were shown on the screen with a visual stimulus presentation system (Presentation, Neurobehavioral Systems, Albany, CA, USA) and a projector (12.5 × 16 × 20″) outside the shielded room. In both experiments, one of the mathematical operation types, i.e., addition or multiplication, was presented on the screen in front of the subject. Then, the subject tries to answer the right answer in his/her mind as instructed upon the presentation of the execution cue. Each trial consisted of the following three phases; these were: (i) the instruction/preparation phase; (ii) the execution phase; and (iii) the rest phase.

In the first experimental paradigm, the participants (17 non-experts and 12 abacus experts) were instructed to perform at least a mental addition of two digits plus two digits' number, immediately after numbers were presented. Confirmation task was prepared to confirm whether the subject performs the task adequately or not but this task was excluded from analysis in this study. We prepared two kinds of tasks in the addition experiment. For the first task, the participants were asked to perform the mental calculation. A black fixation cross (visual fixation condition "+") was presented after each trial for the preparation phase (see **Figure 2**). This digit stimuli "+" was presented for 4 s on the center of a screen (**Figure 1**). The participant was instructed to mentally add the presented series of digits without moving their body, especially, fingers. After the presentation of these digit stimuli "+," series of digits was presented for 2 s. For the second task, the participants were asked to judge whether the addition answer in their mind and the test digit stimuli were the same or different, by answering aloud after each trial. The experimental session consisted of 80 trials for calculation and observation tasks and eight trials for confirmation in a random order.

In the second experimental paradigm, the participants (17 non-experts and 13 abacus experts) were instructed to perform mental multiplication at least by multiplying a two-digit number by a one-digit number (e.g., 72 × 3). We prepared three kinds of tasks in the multiplication experiment (calculation, observation, and confirmation). In the confirmation task, the subjects were asked to speak the answer loudly to confirm whether they perform the given task as requested. This experimental session consisted of 168 trials in total including calculation and confirmation tasks (see **Figure 2**).

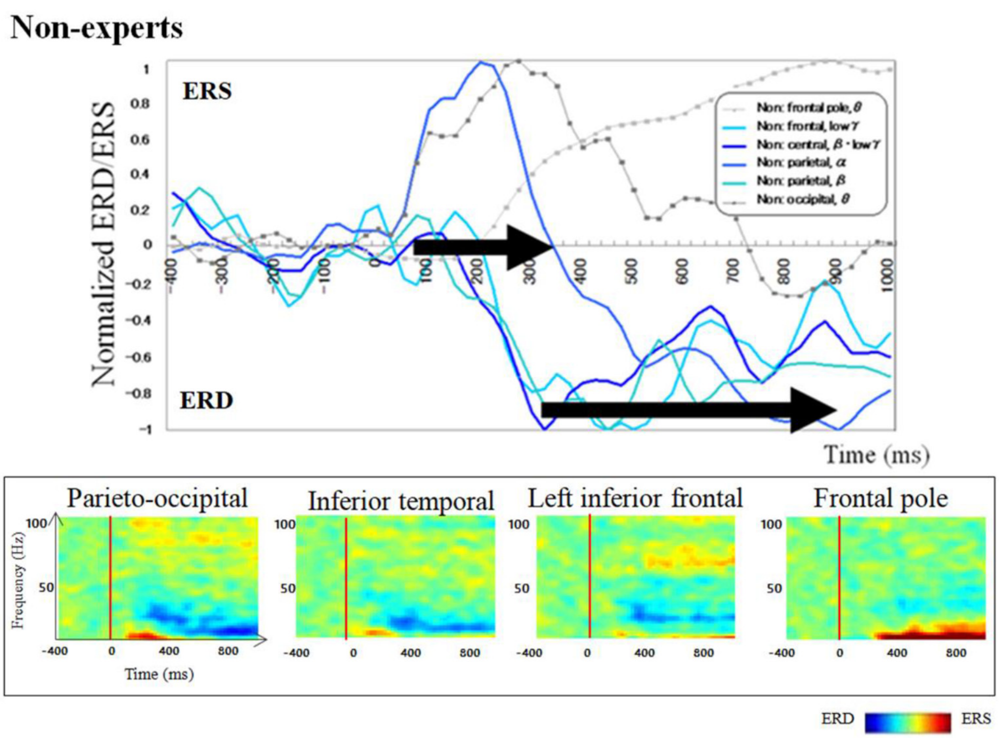


FIGURE 7 | A group average of a time-frequency profile of oscillatory changes for non-experts during mental addition experiments. The upper graph shows the time course of the normalized oscillatory changes in the frequency bands specific to the detected areas and lower graphs show time-frequency spectrograms in each area. The time window is between −400 and 1,000 ms; the movement onset is at 0 ms, and the frequency interval ranges from 4 Hz to 100 Hz. The red areas in the time-frequency maps indicate increases in power (ERS), and the blue areas indicate decreases in power (ERD). In the figure legend, “Non.” means non-abacus experts.

However, we adjusted the task difficulty for both experiments to equalize the task demand for each subject and we used full power metal calculation to detect the calculation-related responses effectively. In each subject, we tested calculation performance (calculation speed and accuracy) in advance before MEG measurements, and equalized task difficulty for each subject. For instance, the subject was instructed to perform 2×1 digits, 2×2 digits, 3×2 digits, 3×3 digits, and 4×3 digits. Each case included 40 calculations. Then, we recorded accuracies of correct answers and times required for calculation. We found no significant difference between experts and non-experts.

Data Acquisition and Pre-processing

The neuromagnetic activity was sampled at 1,000 Hz, after which the temporal extension of the signal space separation method (tSSS) was used to suppress noise and artifacts generated by the sensors and sources of interference located very close to the MEG sensors. A band-pass filter from 0.5 to 100 Hz and a 60-Hz notch filter was applied to reduce some high-frequency noises and eliminate AC line noise.

For analyses, we used a beamforming method called synthetic aperture magnetometry (SAM) with an adaptive beamformer to obtain high spatial resolution. Besides, we introduced a group analysis to exclude the inter-individual variance. The following **Figures 3**, **4** show the time windows of the tasks and analyses used in this study. As for SAM analyses, we compared 1 s before and after number presentation. Moreover, concerning time-frequency analyses, the baseline was set from −400 ms to 0 s. Frequency bands were divided into these five bands; ranging from theta to high gamma bands.

We used synthetic aperture magnetometry (SAM) group analysis; SAM was also applied for MEG analysis using an adaptive beamformer. SAM has high spatial resolution using virtual narrow apertures with which we could detect a high-frequency response because this method does not include the averaging process that otherwise cancels out the high-frequency components. SAM group analysis detects common brain activities from individual SAM images using statistical non-parametric mapping. For group analysis, we used statistical non-parametric mapping (SNMP), which is an option of SPM delivered by the Welcome Department of University College London (UCL). We calculated the non-parametric pseudo-t-statistic images based on the variance-covariances of the voxel-level variances for each frequency band, Family-Wise Error rate (FWE) for $p = 0.001$, and MNI coordinates (X, Y, Z) for Brodmann area and the anatomical localization (see **Supplementary Tables S1–S8**). We also calculated time-frequency analysis using a Morlet wavelet

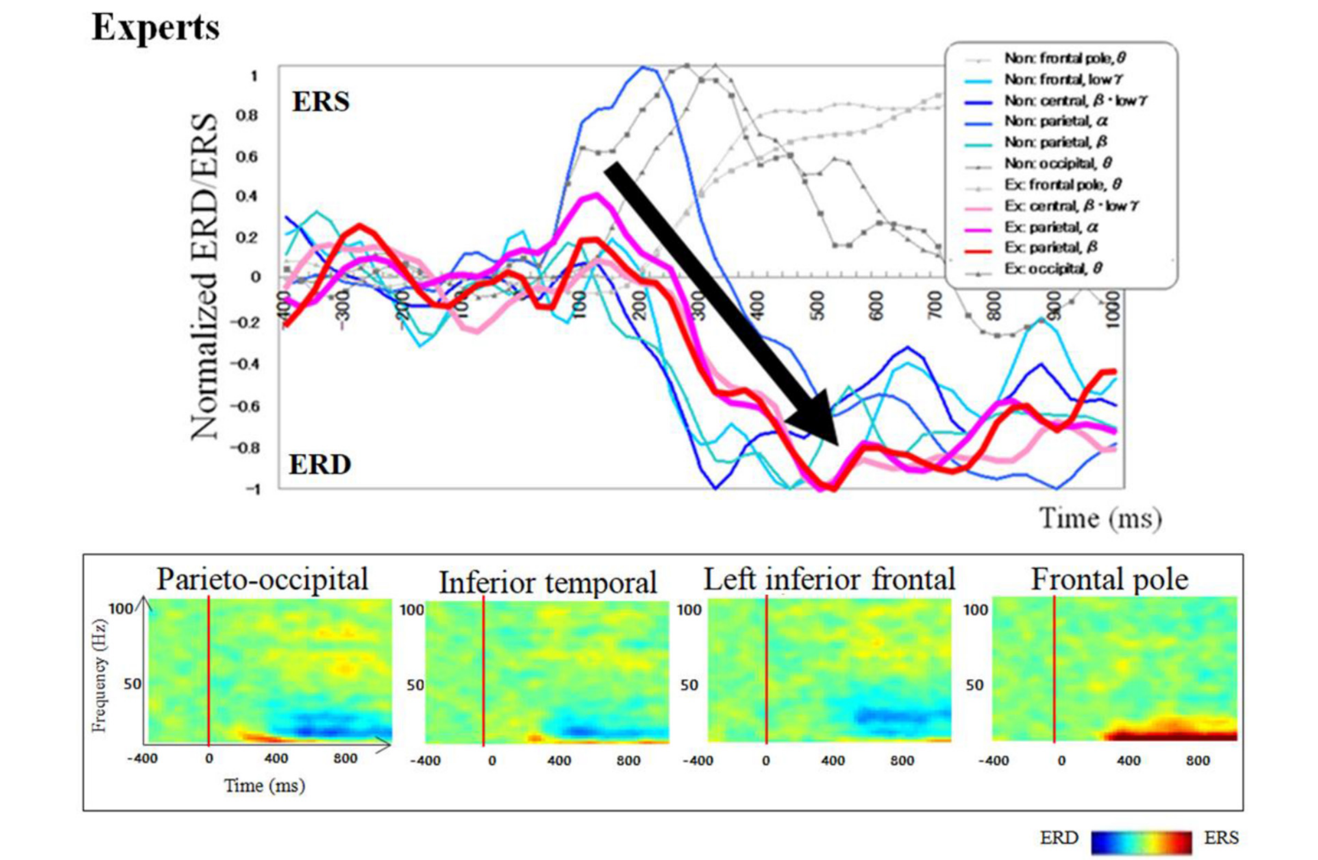


FIGURE 8 | A group average of a time-frequency profile of oscillatory changes for Abacus experts during mental addition experiment. The upper graph shows the time course of the normalized oscillatory changes in the frequency bands specific to the detected areas and lower graphs show time-frequency spectrograms in each area. The time window is between −400 and 1000 ms; the movement onset is at 0 ms, and the frequency interval ranges from 4 Hz to 100 Hz. The red areas in the time-frequency maps indicate increases in power (ERS), and the blue areas indicate decreases in power (ERD). In the figure legend, “Non.” means non-abacus experts and “Ex.” means abacus experts.

transform to elucidate temporal dynamics of the calculation process using MATLAB 2016a software (Mathworks, Natick, MA, USA).

RESULTS

Spatial Distribution of Oscillatory Changes

We analyzed the spatiotemporal frequency patterns of oscillatory changes for experts and non-experts using SAM to understand better source based-brain activation related to abacus experts and investigate calculation-induced neuromagnetic responses based on cerebral oscillatory changes using source level-based MEG signals. The frequency bands are theta (4–8 Hz), alpha (8–13 Hz), beta (13–25 Hz), low gamma (25–50 Hz), high gamma (50–100 Hz). In this study, when subjects performed mental calculations, the magnitude of neuromagnetic fields and the frequency power of brain activities were either increased or decreased in both brain hemispheres; these phenomena are termed as an event-related magnetic field (ERF) for the magnetic fields, event-related synchronization (ERS) and event-related desynchronization (ERD) for the frequency power (Pfurtscheller, 1977, 1992). The spatiotemporal distributions of ERD and ERS could be obtained precisely using SAM. We used this technique to investigate language processing based on cerebral oscillatory changes and have previously reported that cerebral oscillatory changes during silent reading are localized in language-related areas (Hirata et al., 2004, 2007, 2010; Ihara et al., 2003a).

In non-experts, we found: (i) power increase (ERS) in frequency band Theta “θ” in the bilateral frontal pole; (ii) power decrease (ERD) from frequency band Alpha “α” to Beta “β” in the bilateral intraparietal sulcus and the bilateral inferior temporal regions also; (iii) power decrease in low gamma (low γ ERD) in the left inferior and middle frontal gyrus and also in the right dorsolateral prefrontal cortex (DLPFC); and (iv) power increase in high gamma band (high γ ERS) in the bilateral medial occipital regions (for observation and calculation tasks, see **Supplementary Figures S1, S2, S13, S14**).

In abacus experts, we found the spatial and frequency distributions are very similar to those of non-experts, but there are two different points in spatial distribution; these were: (i) in non-experts, α ERD was dominant in the temporal region, but in experts, β ERD was more prominent in the parietal region; and

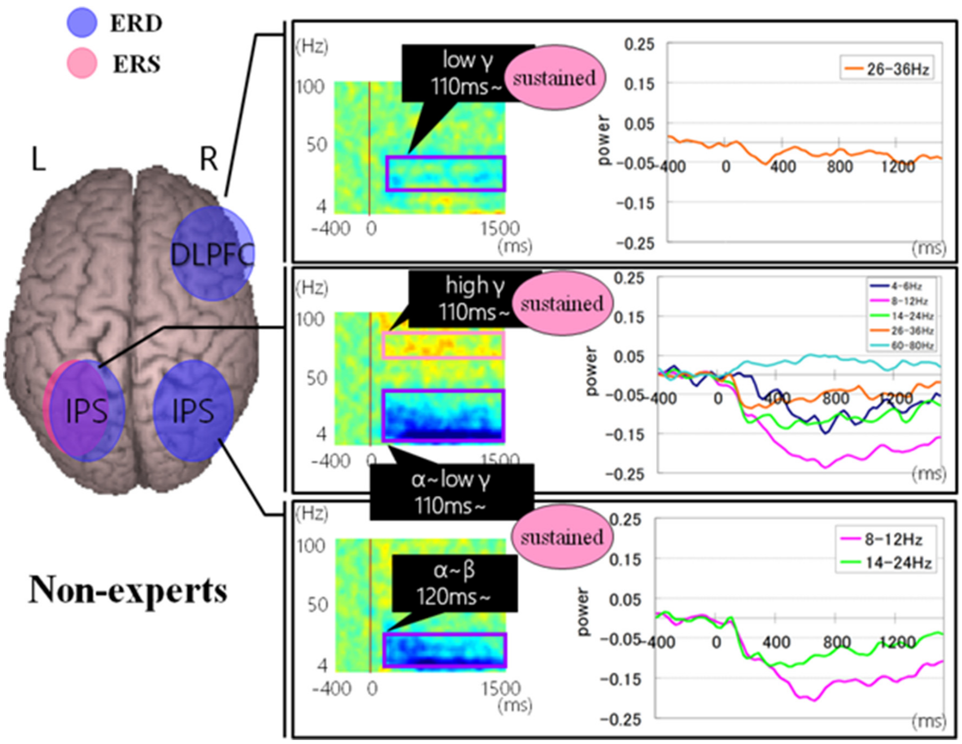


FIGURE 9 | A group average of time-frequency analyses in non-experts during the mental multiplication experiment. The time window is between −400 and 1500 ms; the movement onset is at 0 ms, and the frequency interval ranges from 4 Hz to 100 Hz. The red areas in the time-frequency maps indicate increases in the high gamma power (ERS), and the blue areas indicate decreases in the beta/alpha/low gamma power (ERD).

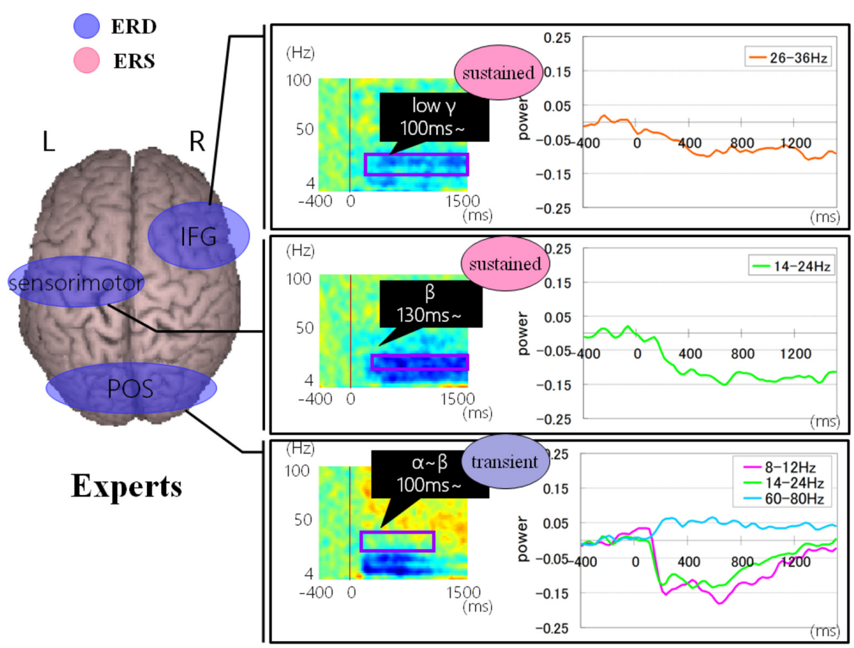


FIGURE 10 | A group average of time-frequency analyses in Abacus experts during the mental multiplication experiment. The time window is between −400 and 1,500 ms; the movement onset is at 0 ms, and the frequency interval ranges from 4 Hz to 100 Hz. The red areas in the time-frequency maps indicate increases in the high gamma power (ERS), and the blue areas indicate decreases in the beta/alpha/low gamma power (ERD).

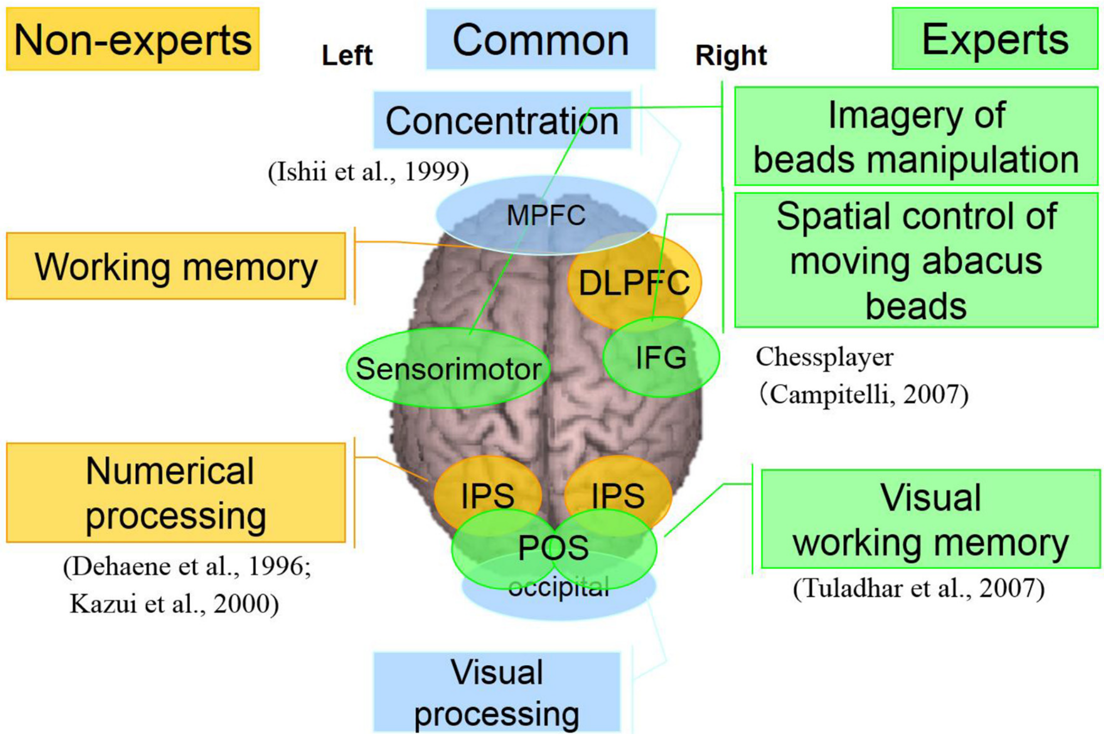


FIGURE 11 | A neural processing mechanism of mental calculation based on cerebral oscillatory changes for Abacus experts and non-experts.

(ii) low γ ERD in the right DLPFC was specific to non-experts only. **Figure 5** shows the spatial distribution of calculation-induced oscillatory changes revealed by SAM group analyses for both abacus experts and novices. The circled areas in these figures indicate statistically significant oscillatory changes (for observation and calculation tasks, see **Supplementary Figures S3, S4, S15, S16**).

Figure 6 shows more activated areas related to mental multiplication tasks. The medial prefrontal cortex and occipital cortex were commonly observed in both experts and non-experts. Specific areas to non-experts are as follows, left dominant parietal mainly intraparietal sulcus (IPS), and right DLPFC. On the other hand, specific areas to abacus experts include bilateral parieto-occipital sulcus (POS), right inferior frontal gyrus (IFG), and left sensorimotor cortex. In the discussion section, we are going to speculate the function of each detected area.

The Temporal Frequency Profile of Oscillatory Changes

Regarding brain activation, we observed some differences in temporal profile for experts and novices using sensor level analysis. We observed, in non-experts, ERDs start serially from parietal, then inferior frontal, dorsolateral prefrontal and finally parietal region again. In contrast, in abacus experts, ERDs start simultaneously in parietal and frontal regions. **Figures 7**, **8** display group average of temporal profile and time-frequency profile of oscillatory changes during the first experimental paradigm while the upper graph shows the time course of the normalized oscillatory changes in the frequency bands specific to the detected areas and lower graphs show time-frequency spectrograms in each area. For observation and calculation tasks, see the spatiotemporal frequency profile of oscillatory changes in **Supplementary Figures S5–S10, S17–S24**.

Figures 9, **10** show group average of time-frequency analyses in abacus experts and non-experts for mental multiplication experiment. For non-experts (see **Figure 9**), the oscillatory responses in the intraparietal sulcus are bilateral, left dominant, and frequency power is sustained. Also, the right DLPFC response was sustained. On the other hand, in experts, interestingly, the parietal response in the bilateral POS is transient. Also, both the right IFG and left sensorimotor responses were sustained (see **Figure 10**).

DISCUSSION

In the present study, we investigated calculation-related oscillatory changes using SAM group analysis and time-frequency analysis. We could elucidate the difference in calculation process between abacus experts and non-experts. In our proposed experimental paradigms, we checked if there may be a drastic difference in the strategy of calculation between abacus experts and non-experts, and

TABLE 1 | Similarities and differences of brain activity during mental operation tasks as revealed by conjoint analysis across subjects in each group for the proposed model of calculation processing.

Regions (most significant Brodmann area) and their function	Non-experts (time-frequency characteristics)		Abacus experts (time-frequency characteristics)	
	Addition	Multiplication	Addition	Multiplication
Medial Occipital for visual processing	✓ θ ERS ~50 ms	✓ θ ERS ~50 ms	✓ θ ERS ~50 ms	✓ θ ERS ~50 ms
The medial prefrontal cortex (mPFC) for concentration (Sasaki et al., 1996; Ishii et al., 1999)	✓ θ ERS ~180 ms	✓ θ ERS ~180 ms	✓ θ ERS ~140 ms	✓ θ ERS ~140 ms
DLPFC for working memory	low γ ERD (Sustained) ~110 ms	low γ ERD (Sustained) ~110 ms	×	×
Right IPS for numerical processing	α, β, and low γ ERD (Sustained) ~110 ms	ERD (Sustained) ~110 ms	×	×
Left IPS for numerical processing	high γ ERS and α, β, and low γ ERD (Sustained) ~110 ms	high γ ERS and α, β, and low γ ERD (Sustained) ~110 ms	×	×
Sensorimotor for beads manipulation of the imaginary abacus	×	×	β ERD (Sustained) ~130 ms	β ERD (Sustained) ~130 ms
IFG for special control of moving abacus beads (Campitelli et al., 2007)	×	×	low γ ERD (Sustained) ~100 ms	low γ ERD (Sustained) ~100 ms
Bilateral POS for visual working memory to transform numbers to abacus beads (Tuladhar et al., 2007)	×	×	α, and β ERD (Transient) ~110 ms	α, and β ERD (Transient) ~110 ms

The symbol "✓" means a common brain area for experts and non-experts. The symbol "×" means that the brain area was not activated during mental calculation.

if abacus experts use an imaginary abacus or abacus memory. Non-experts claim to calculate any visualized numbers using memorized multiplication or addition tables while abacus experts claim that abacus is visualized in the front of their eyes or their mind and the beads move automatically when they try to solve mathematical operations. This study aims to elucidate the difference in neural processing mechanism of mental calculation between experts and non-experts using magnetic source imaging (Della Puppa et al., 2015). Few studies tried to clarify the spatiotemporal characteristics of brain activity during addition and multiplication calculation tasks (Ishii et al., 2014; Vansteensel et al., 2014; Ueda et al., 2015).

Taking into account of temporal profiles of oscillatory changes in the first experimental paradigm, we concluded that non-experts might use serial processing; in contrast, experts may utilize parallel processing (see **Supplementary Figures S11, S12**). We could propose calculation processing in both abacus expert and non- experts based on our new findings and some previous studies (Hanakawa et al., 2002, 2003; Arsalidou and Taylor, 2011; Tanaka et al., 2012; Pinel and Dehaene, 2013; Amalric and Dehaene, 2017). In non-experts, from 75 ms visual processing of presented numbers start in the bilateral medial occipital, then from 150 ms figurative cognition of numbers in the inferior temporal (Dehaene et al., 1996; Pinel et al., 1999) and numeric processing in the bilateral IPS (Dehaene et al., 1996; Cohen et al., 2000; Rickard et al., 2000; Kazui et al., 2000; Bugden et al., 2019), from 200 ms inner speech in the left IFG (Dehaene et al., 1996), from 250 ms working memory in the DLPFC (Rickard et al., 2000), and finally from 400 ms addition with carrying in the IPS starts. Therefore, the calculation in non-experts is serial processing. In contrast, abacus experts calculate using parallel processing, following visual processing from 75 ms, figurative cognition in the inferior temporal, numeric processing in the IPS, and inner speech in the IFG start simultaneously from 250 ms. **Table 1** shows the similarities and differences of brain activities in non-experts and abacus experts during mental operation tasks.

In both experiments where the participants were asked to perform mental operations, we observed some common brain activities in both experts and non-experts. Also, the right DLPFC and bilateral IPS were explicitly detected in non-experts. Bilateral IPS is related to numerical processing (Bugden et al., 2019), while the right DLPFC is most probably related to working memory. In experts, bilateral POS, right IFG, and left sensorimotor areas were detected specifically. POS is related to visual working memory. This response was transient, so probably used merely to transform numbers to abacus beads rather than numerical processing. Right IFG activation is reported in a Chess-players' study (Campitelli et al., 2007). The study suggested IFG plays a crucial role in working memory of strategic spatial configuration of chess pieces. So in the present study, the right IFG may play a vital role in the strategic spatial control of moving abacus beads. The left sensorimotor area is probably related to the imagery of beads manipulation. After long time training, these areas might act as an imaginary abacus or abacus memory, so that experts do not have to logical numerical processing in left IPS, instead quickly perform complex calculation just by retrieving memorized abacus memory as if spatial pattern matching (see **Supplementary Figure S25**). Finally, for our proposed model of calculation processing in normal people and abacus experts, we do believe that normal people calculate logically depending on numerical processing in left IPS. In contrast, abacus experts utilize spatial processing using memorized imaginary abacus, which distributed over the bilateral hemispheres (see **Figure 11**).

As noted above, the present Abacus-based mental calculation experiments sought to provide a conceptual advancement. Notably, being able to find a model of calculation processing, the calculation-related area is an essential step toward understanding the brain processing mechanism, thereby leading to enhance the mathematical performance of patients with Acalculia. Although further studies are required, we would like to apply these findings (see **Table 1**) clinically to the less invasive neurosurgical treatments, considering not only standard calculation-related areas but also inter-individual variation including extreme brains.

Another important question still unanswered refers to the challenges of performing extremely complex calculations (5 × 5 digits). In this study, the experts have high-level mental calculation with 5–27 years' experience, but we did not include world top-level experts. These top-level experts may show completely different brain activities, which means different neural mechanisms during the complex mental calculation. However, these experts are not likely to be able to stop mental calculation even during observation tasks because the results will automatically come to their minds instantly. It means that these top-level experts may perform mental calculations automatically independent from their will and this unanswered question should be addressed in the future.

CONCLUSION

For the abacus experts in performing computations, different brain areas are involving in beads manipulation, and special control of imaginary abacus was observed. These unique findings suggested that, through an effective abacus training, the experts developed a new computational pathway by assigning number representations onto an imaginative abacus representation, through a different brain network.

We concluded that non-experts might use serial processing; in contrast, experts may utilize parallel processing. Abacus experts may acquire this processing system after long time training, and their MEG results demonstrated that calculation-related areas are simultaneously activated using abacus within the brain, and these parallel processing processes indeed significantly shorten the computation time.

DATA AVAILABILITY STATEMENT

The datasets generated and analyzed during the current study are not publicly available due to Osaka University Hospital policy but are available from the corresponding author on reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Committee at Osaka University Hospital. The experimental protocol was carried out according to the latest version of the Declaration of Helsinki. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MH designed the study. KK and EU performed experiments and analyses. AB performed analyses, literature review, and drafted the manuscript. MH supervised the research and revised the manuscript. All authors provided critical feedback, reviewed, edited and approved the final version of the manuscript.

FUNDING

This work was funded partially by the Endowed Research Fund of the Endowed Research Department of Clinical Neuroengineering, and KAKENHI (23590672, 23390347). Grant funded by the Japan Society for the Promotion of Science (JSPS).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00137/full#supplementary-material.

REFERENCES

Amalric, M., and Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proc. Natl. Acad. Sci. U S A* 113, 4909–4917. doi: 10.1073/pnas.1603205113

Amalric, M., and Dehaene, S. (2017). Cortical circuits for mathematical knowledge: evidence for a major subdivision within the brain's semantic networks. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 373:20160515. doi: 10.1098/rstb.2016.0515

Amalric, M., and Dehaene, S. (2019). A distinct cortical network for mathematical knowledge in the human brain. *NeuroImage* 189, 19–31. doi: 10.1016/j.neuroimage.2019.01.001

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

Bugden, S., Woldorff, M. G., and Brannon, E. M. (2019). Shared and distinct neural circuitry for nonsymbolic and symbolic double-digit addition. *Hum. Brain Mapp.* 40, 1328–1343. doi: 10.1002/hbm.24452

Campitelli, G., Gobet, F., Head, K., Buckley, M., and Parker, A. (2007). Brain localization of memory chunks in chessplayers. *Int. J. Neurosci.* 117, 1641–1659. doi: 10.1080/00207450601041955

Chen, F., Hu, Z., Zhao, X., Wang, R., Yang, Z., Wang, X., et al. (2006). Neural correlates of serial abacus mental calculation in children: a functional MRI study. *Neurosci. Lett.* 403, 46–51. doi: 10.1016/j.neulet.2006.04.041

Cohen, L., Dehaene, S., Chochon, F., Lehéricy, S., and Naccache, L. (2000). Language and calculation within the parietal lobe: a combined cognitive, anatomical and fMRI study. *Neuropsychologia* 38, 1426–1440. doi: 10.1016/s0028-3932(00)00038-5

Cohen Kadosh, R., Cohen Kadosh, K., Kaas, A., Henrik, A., and Goebel, R. (2007). Notation-dependent and -independent representations of the number in the parietal lobes. *Neuron* 53, 307–314. doi: 10.1016/j.neuron.2006.12.025

Cohen Kadosh, R., Lammertyn, J., and Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Prog. Neurobiol.* 84, 132–147. doi: 10.1016/j.pneurobio.2007.11.001

Cohen, L., Jobert, A., Le Bihan, D., and Dehaene, S. (2004). Distinct unimodal and multimodal regions for word processing in the left temporal cortex. *NeuroImage* 23, 1256–1270. doi: 10.1016/j.neuroimage.2004.07.052

Cohen, J., Perlstein, W., Braver, T., Nystrom, L., Noll, D., Jonides, J., et al. (1997). Temporal dynamics of brain activation during a working memory task. *Nature* 386, 604–608. doi: 10.1038/386604a0

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-n

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

Dehaene, S. (1996). The organization of brain activity in number comparison: event-related potentials and the additive-factors methods. *J. Cogn. Neurosci.* 8, 47–68. doi: 10.1162/jocn.1996.8.1.47

Dehaene, S. (1997). *The Number Sense.* Oxford: Oxford University Press.

Dehaene, S. (2008). "Symbols and quantities in parietal cortex: elements of a mathematical theory of number representation and manipulation," in *Attention and Performance XXII*, eds P. Haggard, Y. Rossetti and M. Kawato (London: Oxford University Press), 527–574.

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., Dupoux, E., and Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in 2-digit number comparison. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 626–641. doi: 10.1037/0096-1523.16.3.626

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., Spelke, E., Pinel, P., Stanescu, R., and Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* 284, 970–974. doi: 10.1126/science.284.5416.970

Della Puppa, A., De Pellegrin, S., Lazzarini, A., Gioffrè, G., Rustemi, O., Cagnin, A., et al. (2015). Subcortical mapping of calculation processing in the right parietal lobe. *J. Neurosurg.* 122, 1038–1041. doi: 10.3171/2014.10.jns14261

Dimitriadis, S. I., Sun, Y., Thakor, N. V., and Bezerianos, A. (2016). Causal interactions between frontal$^{\theta}$—parieto-occipital$^{\alpha}$ predict performance on a mental arithmetic Task. *Front. Hum. Neurosci.* 10:454. doi: 10.3389/fnhum.2016.00454

Hanakawa, T., Honda, M., Okada, T., Fukuyama, H., and Shibasaki, H. (2003). Neural correlates underlying mental calculation in abacus experts: a functional magnetic resonance imaging study. *NeuroImage* 19, 296–307. doi: 10.1016/s1053-8119(03)00050-8

Hanakawa, T., Honda, M., Sawamoto, N., Okada, T., Yonekura, Y., Fukuyama, H., et al. (2002). The role of rostral Brodmann area6 in mental-operation tasks: an integrative neuroimaging approach. *Cereb. Cortex* 12, 1157–1170. doi: 10.1093/cercor/12.11.1157

Hirata, M., Goto, T., Barnes, G., Umekawa, Y., Yanagisawa, T., Kato, A., et al. (2010). Language dominance and mapping based on neuromagnetic oscillatory changes: comparison with invasive procedures. *J. Neurosurg.* 112, 528–538. doi: 10.3171/2009.7.jns09239

Hirata, M., Kato, A., Taniguchi, M., Saitoh, Y., Ninomiya, H., Ihara, A., et al. (2004). Determination of language dominance with synthetic aperture magnetometry: comparison with the Wada test. *NeuroImage* 23, 46–53. doi: 10.1016/j.neuroimage.2004.05.009

Hirata, M., Koreeda, S., Sakihara, K., Kato, A., Yoshimine, T., and Yorifuji, S. (2007). Effects of the emotional connotations in words on the frontal areas—a spatially filtered MEGstudy. *NeuroImage* 35, 420–429. doi: 10.1016/j.neuroimage.2006.11.025

Hu, Y., Geng, F., Tao, L., Hu, N., Du, F., Fu, K., et al. (2011). Enhanced white matter tracts integrity in children with abacus training. *Hum. Brain Mapp.* 32, 10–21. doi: 10.1002/hbm.20996

Iguchi, Y., and Hashimoto, I. (2000). Sequential information processing during a mental arithmetic is reflected in the time course of event-related brain potentials. *Clin. Neurophysiol.* 111, 204–213. doi: 10.1016/s1388-2457(99)00244-8

Ihara, A., Hirata, M., Sakihara, K., Izumi, H., Takahashi, Y., Kono, K., et al. (2003a). Gamma-banddesynchro-nation in language areas reflects the syntactic process of words. *Neurosci. Lett.* 339, 135–138. doi: 10.1016/s0304-3940(03)00005-3

Ihara, A., Hirata, M., Yanagihara, K., Ninomiya, H., Imai, K., Ishii, R., et al. (2003b). Neuromagneticgamma-band activity in the primary and secondary somatosensory areas. *Neuroreport* 14, 273–277. doi: 10.1097/00001756-200302100-00024

Ishii, R., Canuet, L., Ishihara, T., Aoki, Y., Ikeda, S., Hata, M., et al. (2014). Frontal midline theta rhythm and gamma power changes during focused attention on the mental calculation: a MEG beamformer analysis. *Front. Hum. Neurosci.* 8:406. doi: 10.3389/fnhum.2014.00406

Kazui, H., Kitagaki, H., and Mori, E. (2000). Cortical activation during retrieval of arithmetical facts and actual calculation: a functional magnetic resonance imaging study. *Psychiatry Clin. Neurosci.* 54, 479–485. doi: 10.1046/j.1440-1819.2000.00739.x

Ishii, R., Shinozaki, K., Ukai, S., Inouye, T., Ishihara, T., Yoshimine, T., et al. (1999). Medial prefrontal cortex generate frontal midline theta rhythm. *Neuroreport* 10, 675–679. doi: 10.1097/00001756-199903170-00003

Li, Y., Hu, Y., Zhao, M., Wang, Y., Huang, J., and Chen, F. (2013). The neural pathway underlying a numerical working memory task in abacus-trained children and associated functional connectivity in the resting brain. *Brain Res.* 1539, 24–33. doi: 10.1016/j.brainres.2013.09.030

Pauli, P., Lutzenberger, W., Rau, H., Birbaumer, N., Rickard, T. C., Yaroush, R. A., et al. (1994). Brain potentials during mental arithmetic: effects of extensive practice and problem difficulty. *Cogn. Brain Res.* 2, 21–29. doi: 10.1016/0926-6410(94)90017-5

Pfurtscheller, G. (1977). Graphical display and statistical evaluation of event-related desynchronization (ERD). *Electroencephalogr. Clin. Neurophysiol.* 43, 757–760. doi: 10.1016/0013-4694(77)90092-x

Pfurtscheller, G. (1992). Event-related synchronization (ERS): an electrophysiological correlate of cortical areas at rest. *Electroencephalogr. Clin. Neurophysiol.* 83, 62–69. doi: 10.1016/0013-4694(92)90133-3

Pinel, P., and Dehaene, S. (2013). Genetic and environmental contributions to brain activation during calculation. *NeuroImage* 81, 306–316. doi: 10.1016/j.neuroimage.2013.04.118

Pinheiro-Chagas, P., Piazza, M., and Dehaene, S. (2018). Decoding the processing stages of mental arithmetic with magnetoencephalography. *Cortex* 114, 124–139. doi: 10.1016/j.cortex.2018.07.018

Pinel, P., Le Clec'H, G., van de Moortele, P. F., Naccache, L., Le Bihan, 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

Rickard, T. C., Romero, S. G., Basso, G., Wharton, C., Flitman, S., and Grafman, J. (2000). The calculating brain: an fMRI study. *Neuropsychologia* 38, 325–335. doi: 10.1016/s0028-3932(99)00068-8

Robinson, S. E., Nagarajan, S. S., Mantle, M., Gibbons, V., and Kirsch, H. (2004). Localization of interictal spikes using SAM(g2) and dipole fit. *Neurol. Clin. Neurophysiol.* 2004:74.

Sasaki, K., Tsujimoto, T., Nishikawa, S., Nishitani, N., and Ishihara, T. (1996). Frontal mental theta wave recorded simultaneously with magnetoencephalography and electroencephalography. *Neurosci. Res.* 26, 79–81. doi: 10.1016/0168-0102(96)01082-6

Tanaka, S., Michimata, C., Kaminaga, T., Honda, M., and Sadato, N. (2002). Superior digit memory of abacus experts: an event-related functional MRI study. *Neuroreport* 13, 2187–2191. doi: 10.1097/00001756-200212030-00005

Tanaka, S., Seki, K., Hanakawa, T., Harada, M., Sugawara, S. K., Sadato, N., et al. (2012). Abacus in the brain: a longitudinal functional MRI study of a skilled abacus user with a right hemispheric lesion. *Front. Psychol.* 3:315. doi: 10.3389/fpsyg.2012.00315

Taniguchi, M., Kato, A., Fujita, N., Hirata, M., Tanaka, H., Kihara, T., et al. (2000). Movement-related desynchronization of the cerebral cortex studied with spatially filtered magnetoencephalography. *NeuroImage* 12, 298–306. doi: 10.1006/nimg.2000.0611

Tuladhar, A. M., ter Huurne, N., Schoffelen, J. M., Maris, E., Oostenveld, R., and Jensen, O. (2007). Parieto-occipital sources account for the increase in alpha activity with working memory load. *Hum. Brain Mapp.* 28, 785–792. doi: 10.1002/hbm.20306

Ueda, K., Brown, E. C., Kojima, K., Juhasz, C., and Asano, E. (2015). Mapping mental calculation systems with electrocorticography. *Clin. Neurophysiol.* 126, 39–46. doi: 10.1016/j.clinph.2014.04.015

Vansteensel, M. J., Bleichner, M. G., Freudenburg, Z. V., Hermes, D., Aarnoutse, E. J., Leijten, F. S., et al. (2014). Spatiotemporal characteristics of electrocortical brain activity during mental calculation. *Hum. Brain Mapp.* 35, 5903–5920. doi: 10.1002/hbm.22593

Wang, Y., Geng, F., Hu, Y., Du, F., and Chen, F. (2013). Numerical processing efficiency improved inexperienced mental abacus children. *Cognition* 127, 149–158. doi: 10.1016/j.cognition.2012.12.004

Wang, C., Weng, J., Yao, Y., Dong, S., Liu, Y., and Chen, F. (2017). Effect of abacus training on executive function development and underlying neural correlates in Chinese children. *Hum. Brain Mapp.* 38, 5234–5249. doi: 10.1002/hbm.23728

Wu, T. H., Chen, C. L., Huang, Y. H., Liu, R. S., Hsieh, J. C., and Lee, J. J. (2009). Effects of long-term practice and task complexity on brain activities when performing abacus-based mental calculations: a PET study. *Eur. J. Nucl. Med. Mol. Imaging* 36, 436–445. doi: 10.1007/s00259-008-0949-0

Conflict of Interest: 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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 May 2020
doi: 10.3389/fpsyg.2020.00871

Hierarchical Development of Early Visual-Spatial Abilities – A Taxonomy Based Assessment Using the MaGrid App

Stefanie Jung[1,2*†], Anna Meinhardt[1,3,4†], David Braeuning[1,5,6], Stephanie Roesch[1], Véronique Cornu[7], Tahereh Pazouki[7], Christine Schiltz[8], Jan Lonnemann[9] and Korbinian Moeller[1,2,4,5,10]

[1] Leibniz-Institut für Wissensmedien, Tübingen, Germany, [2] Department of Psychology, Eberhard Karls University Tübingen, Tübingen, Germany, [3] DIPF Leibniz Institute for Research and Information in Education, Frankfurt, Germany, [4] Center for Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt, Germany, [5] LEAD Graduate School and Research Network, University of Tübingen, Tübingen, Germany, [6] Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Tübingen, Germany, [7] Luxembourg Centre for Educational Testing (LUCET), University of Luxembourg, Esch-sur-Alzette, Luxembourg, [8] Department of Behavioral and Cognitive Sciences (DBCS), University of Luxembourg, Esch-sur-Alzette, Luxembourg, [9] Empirical Childhood Research, University of Potsdam, Potsdam, Germany, [10] Centre for Mathematical Cognition, School of Science, Loughborough University, Loughborough, United Kingdom

OPEN ACCESS

Edited by:
*David Peebles,
University of Huddersfield,
United Kingdom*

Reviewed by:
*Corinne Bower,
University of Maryland, College Park,
United States
Ilyse Resnick,
University of Canberra, Australia*

***Correspondence:**
*Stefanie Jung
s.jung@iwm-tuebingen.de*

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *12 December 2019*
Accepted: *07 April 2020*
Published: *20 May 2020*

Citation:
Jung S, Meinhardt A, Braeuning D, Roesch S, Cornu V, Pazouki T, Schiltz C, Lonnemann J and Moeller K (2020) Hierarchical Development of Early Visual-Spatial Abilities – A Taxonomy Based Assessment Using the MaGrid App. Front. Psychol. 11:871. doi: 10.3389/fpsyg.2020.00871

Visual-spatial abilities (VSA) are considered a building block of early numerical development. They are intuitively acquired in early childhood and differentiate in further development. However, when children enter school, there already are considerable individual differences in children's visual-spatial and numerical abilities. To better understand this diversity, it is necessary to empirically evaluate the development as well as the latent structure of early VSA as proposed by the 2 by 2 taxonomy of Newcombe and Shipley (2015). In the present study, we report on a tablet-based assessment of VSA using the digital application (app) MaGrid in kindergarten children aged 4–6 years. We investigated whether the visual-spatial tasks implemented in MaGrid are sensitive to replicate previously observed age differences in VSA and thus a hierarchical development of VSA. Additionally, we evaluated whether the selected tasks conform to the taxonomy of VSA by Newcombe and Shipley (2015) applying a confirmatory factor analysis (CFA) approach. Our results indicated that the hierarchical development of VSA can be measured using MaGrid. Furthermore, the CFA substantiated the hypothesized factor structure of VSA in line with the dimensions proposed in the taxonomy of Newcombe and Shipley (2015). Taken together, the present results advance our knowledge to the (hierarchical) development as well as the latent structure of early VSA in kindergarten children.

Keywords: visual-spatial abilities, 2 by 2 taxonomy, geometry, tablet-based approach, MaGrid

INTRODUCTION

Early numerical development was suggested to build on both spatial-geometric and numerical-quantitative concepts and the acquisition of corresponding abilities (Sarama and Clements, 2004; Jirout and Newcombe, 2015; Newcombe et al., 2015). These skills were argued to be acquired intuitively in early childhood (e.g., Newcombe et al., 2015), but their close association persists in adulthood (Dehaene et al., 1999; Hubbard et al., 2005).

However, already at the age of kindergarten, there are large individual differences in children's spatial and numerical skills (Krajewski and Schneider, 2009; Newcombe and Frick, 2010), which also have long-term consequences: For example, longitudinal studies revealed that children's spatial as well as basic numerical abilities at the age of kindergarten predict their mathematical achievement in primary school and beyond (Duncan et al., 2007; Krajewski and Schneider, 2009; Verdine et al., 2017). More recent evidence from large-scale factor analytic studies suggested strong relations among visual-spatial and mathematic skills in first, third and sixth graders (Mix et al., 2016, 2017).

Visual-spatial abilities (VSA), in particular, are an important building block when it comes to acquiring geometric abilities (Franke and Reinhold, 2007), indicating that their impact goes beyond typically considered basic numerical abilities such as counting and magnitude understanding (cf. Clements, 1998). However, there are multiple abilities summarized under the broad umbrella of VSA for which it is difficult to specify theoretical concepts associated with this term (Eliot and Smith, 1983; Carroll, 1993; Newcombe et al., 2015; Mix et al., 2016). Only recently, Newcombe and Shipley (2015) proposed a top-down systematic taxonomy of VSA, which considers and integrates prior distinctions of different dimensions of VSA. This taxonomy defines VSA along two dimensions: first, VSA being either intrinsic to vs. extrinsic between objects (following the neural organization of spatial thinking, e.g., Chatterjee, 2008). Second, VSA being related to static vs. dynamic aspects of objects (considering propositions by e.g., Kozhevnikov et al., 2002). Such a systematic attempt to define the actual nature of VSA and to understand their latent cognitive components may provide a promising framework based on which VSA can be assessed and promoted.

In the present paper, we aimed at validating the 2 by 2 taxonomy of Newcombe and Shipley (2015) using an assessment procedure for VSA in kindergarten children aged 4 to 6 years from both a theoretical and a behavioral perspective. From a theoretical perspective, we investigated how VSA develop with respect to the intrinsic-extrinsic dimension as well as to the static-dynamic dimension as proposed in the 2 by 2 taxonomy of VSA. From a behavioral perspective, we investigated the hierarchical development of VSA as assessed by the digital application (app) MaGrid ("Math on Grid"; Cornu et al., 2017; Pazouki et al., 2018). In the following, we will first report on recent approaches to theoretically categorize VSA before we consider their hierarchical development. Subsequently, we introduce the tablet-based app MaGrid to provide an idea of its functionality and how the app is currently used to promote VSA.

A Taxonomy of Visual-Spatial Abilities

A comprehensive understanding of VSA, which are generally referred "to skill[s] in representing, transforming, generating, and recalling symbolic, non-linguistic information" (Linn and Petersen, 1985, p. 1,482), is essential to develop valid assessment and training tools. However, its complexity has long hampered a coherent definition. Still today, there are inconsistencies and contradictions in the literature on VSA. Although different bottom-up factor-analytical approaches have confirmed the variety of spatial abilities (Newcombe and Shipley, 2015), they did not lead to a consensus on the definition of this term.

Uttal et al. (2013) were among the first to adopt an opposing top-down approach: they worked on the development of a two-dimensional classification system of VSA. This classification system is referred to by the 2 by 2 taxonomy proposed by Newcombe and Shipley (2015) and incorporates evidence from cognitive, linguistic and neural findings (Palmer, 1978; Talmy, 2000; Chatterjee, 2008). Within this taxonomy, four different categories of VSA are defined: Intrinsic-static (i.e., perceiving objects), intrinsic-dynamic (i.e., assembling small units into larger ones, mental rotation), extrinsic-static (i.e., understanding abstract spatial concepts), and extrinsic-dynamic (i.e., perspective taking) VSA.

Intrinsic processes require only consideration of the object at hand, whereas object surroundings in terms of a reference frame are not considered. A reference frame is understood as a coordinate system needed to determine the position of an object in space in relation to others from a certain perspective (Talmy, 2000). Extrinsic processes, in contrast, involve relations between different objects as well as the spatial configuration of objects within a reference frame. Static and dynamic aspects of single or multiple objects concern the immobility or motion of objects. On the one hand, an object can remain static, which means that it does not change its position, orientation, and/or dimension. On the other hand, objects can be manipulated physically or mentally, which involves changes in position and orientation. This manipulation defines dynamic VSA. For example, the picture of a car can be viewed as a 2D-static object. The car itself, however, can also be viewed as a 3D dynamic object. In 3D, the car can be rotated or moved. It is also possible to take, for instance, the perspective of its driver.

Literature on VSA provides considerable support for the 2 by 2 taxonomy of Newcombe and Shipley (2015, e.g., Newcombe, 2018, for a review). It is therefore increasingly used as a theoretical framework for the classification of VSA. For example, Hodgkiss et al. (2018) tested VSA of 7- to 11-years-old children using five different tasks, which the authors assigned to the four categories of VSA according to the 2 by 2 taxonomy (i.e., intrinsic-static: visual embedding; intrinsic-dynamic: mental rotation and mental folding; extrinsic-static: spatial scaling; extrinsic-dynamic: photo spatial perspective taking). They observed that task performance differed significantly between categories. Interestingly, only intrinsic-dynamic and extrinsic-static VSA were found to predict performance in STEM subjects (e.g., biology, chemistry, physics). However, while this provides evidence corroborating the taxonomy of Newcombe and Shipley (2015) the findings of Hodgkiss et al. (2018) do not yet reflect a validation of the taxonomy. To do so, it would be necessary to include more than one task per category of VSA and to evaluate the relations within vs. between tasks and categories, which the authors did only for intrinsic-dynamic VSA.

In contrast, Mix et al. (2018) assessed two tasks per category of the 2 by 2 taxonomy in a *post hoc* analysis of previously published data (Mix et al., 2017). However, their findings did not support the validity of the theoretically assumed 2 by 2 structure of VSA.

Using a confirmatory factor analysis (CFA) approach on data of school children (i.e., first, third and sixth grade), the authors did not observe evidence for an overall 2 by 2 structure. Instead, their CFA results showed that the static-dynamic 2-factor model did not provide a better fit than a single factor model. Consequently, there was no differentiation along the static-dynamic dimension of VSA. Furthermore, the differentiation between intrinsic and extrinsic VSA was substantiated by the CFA, but only for first and third graders. For sixth graders, a single factor model was found to fit the data best. Based on these findings, Mix et al. (2018) suggested that the latent structure of VSA may change over the course of their development. They proposed to further investigate the developmental trajectories of VSA which was one aim of the present study.

Hierarchical Development of VSA Considering the 2 by 2 Taxonomy

Studies on the early development of VSA demonstrated that these abilities begin to develop already in infancy and further evolve during childhood (Frick and Wang, 2014). From the literature, it is reasonable to assume that this hierarchical development of VSA may also be reflected in the 2 by 2 taxonomy of Newcombe and Shipley (2015) although the complexity involved in categorizing VSA and tasks can hardly be captured by such an approach (Newcombe, 2018).

In the course of development, it is assumed that the development of intrinsic VSA precedes the development of extrinsic VSA (Newcombe and Huttenlocher, 2006). Similarly, the development of static VSA is assumed to precede the development of dynamic VSA (Okamoto et al., 2015). In particular and concerning the intrinsic-static category, Clements (1998), for example, analyzed the characteristics by which 3–6 years old kindergarten children distinguish between different shapes (e.g., circles and rectangles). The authors observed that almost all children were able to recognize and externally verbalize the object's characteristics. However, they also found that object recognition did improve with age.

Similar results were reported by Stiles and Tada (1996) who assessed how kindergarten children of different age groups (i.e., 3–3.5, 3.5–4, 4–4.5, and 4.5–5 years) segmented objects (e.g., +, ×, *) into parts or integrated parts to objects. The authors found that younger children segmented forms into more components than older children, because they perceived lines as discontinuous due to, for instance, an intersection at the midpoint. Older children, instead, perceived the lines as continuous across such an intersection. This indicates that they already seem to have acquired more elaborate shape recognition skills and thus a more abstract representation of the respective object.

Based on such an abstract representation of forms and objects (e.g., length and distance of lines, or angles; Lee et al., 2012), children may then develop extrinsic-static abilities that involve an understanding of spatial relations between objects and the environment as well as the size and scaling of objects. Then again, processing of extrinsic-static information improves with age and individual experiences (Newcombe and Huttenlocher, 2006; see also Okamoto et al., 2015, for an overview).

In contrast to the understanding of intrinsic-static or extrinsic-static characteristics of objects, dynamic VSA often involve transforming, (mentally) rotating, or assembling (a set of) objects as well as perspective taking (Uttal et al., 2013; Newcombe and Shipley, 2015). With regard to the intrinsic-dynamic category, Clements et al. (2004) investigated the development of this VSA in 3–7 years old children in a composition task of geometric figures. The successful development of intrinsic-dynamic VSA is seen as a prerequisite to cope with extrinsic-dynamic visual-spatial processing because extrinsic-dynamic VSA involve recognition of changing spatial relations of objects while considering the environment from different perspectives. Thereby, they involve self-to-object (i.e., perspective taking) and object-to-object (i.e., location learning) navigation (Okamoto et al., 2015), which develop throughout the early years of childhood.

Despite the consideration of the different dimensions of VSA, the development of VSA along the intrinsic-extrinsic dimension cannot be assumed to be distinct from the development of VSA along the static-dynamic dimension. More likely, a development across both dimensions can be assumed. To be more specific, when considering the four categories of VSA as a 2 × 2 matrix (see also **Figure 1**), developmental trajectories would be expected both in the horizontal direction along the static-dynamic dimension as well as in the vertical direction along the intrinsic-extrinsic dimension. Consequently, intrinsic-static VSA are assumed to develop earlier than intrinsic-dynamic VSA while they also develop earlier than extrinsic-static VSA. Accordingly, within a specific age group, intrinsic-static VSA should be further developed than intrinsic-dynamic VSA, which should be more pronounced than extrinsic-static VSA and these again further developed than extrinsic-dynamic VSA (i.e., intrinsic-static > intrinsic-dynamic > extrinsic-static > extrinsic-dynamic). Based on this assumption, the 2 × 2 taxonomy of VSA by Newcombe and Shipley (2015) provides a framework not only for the structure of VSA but also for the development of VSA with age (Uttal et al., 2013; for the malleability of VSA).

Latest developments in digital technologies are influencing the development of assessment and training tools for VSA at an incredible speed, providing small and ready to use devices such as touch-operated smartphones and tablet devices. Tablets, in particular, are increasingly used in educational settings (e.g., Goodwin, 2012; Pazouki et al., 2018). Tablet-based trainings have been shown to improve VSA – even though these improvements have been found to differ from improvements gained in paper-based trainings (e.g., Lowrie et al., 2014, but see Lowrie et al., 2017), for partly contradictory results).

There is, however, no requirement for scientific validation for apps marketed as educational (Hirsh-Pasek et al., 2015). This is problematic for educators and parents alike when they want to ensure that children are using appropriate and effective apps for educational purposes (Hirsh-Pasek et al., 2015). In turn, this emphasizes the need for research and development of validated educational apps.

From the perspective of an app, tablets already seem to be attractive to young children as they encourage kindergarten children to become more closely and effectively involved in

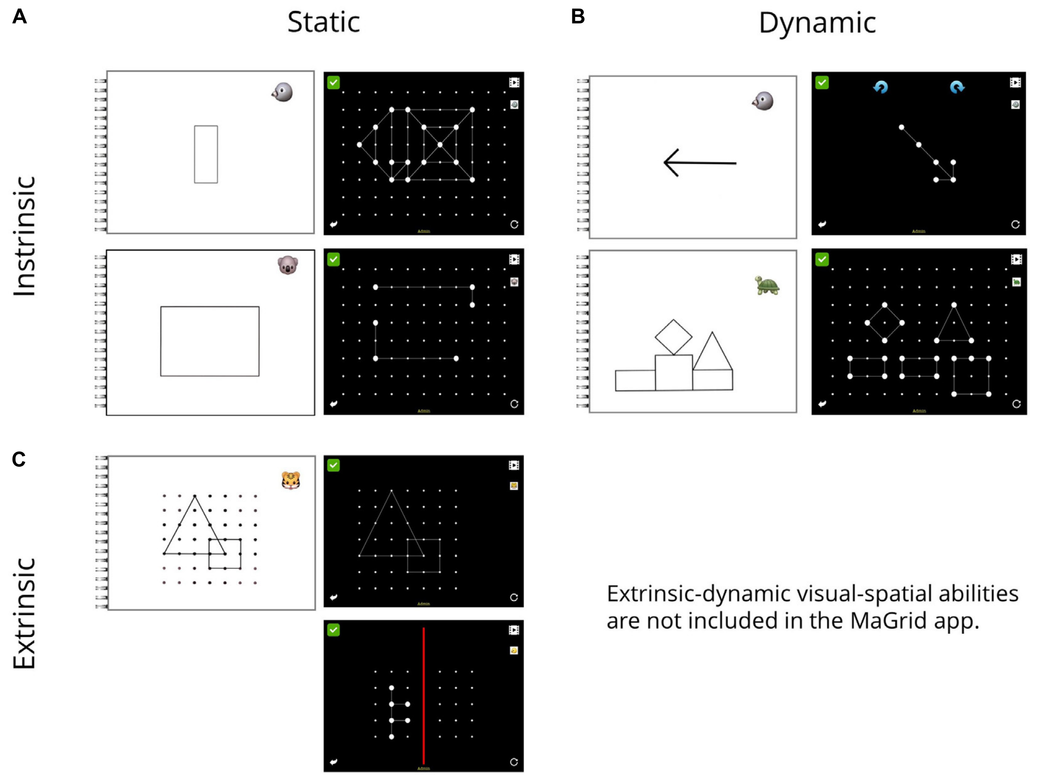


FIGURE 1 | (A–C) Classification of MaGrid tasks in the taxonomy proposed by Newcombe and Shipley (2015). **(A)** Depicts an example of an item of the *Find forms* task in the top row and an item of the *Close form* task in the bottom row. For both tasks, the tablet display is shown on the right side and the corresponding booklet picture on the left side of the panel. **(B)** Gives an example of an item of the *Rotation* task in the top row and an item of the *Tangram* task on the bottom row. For both tasks, target objects are given in the booklet (left). **(C)** Shows an item of the *Reproduce forms I* task for which again a booklet is needed (left). The bottom row depicts an example of the *Reproduce forms II* task. Only the tablet is needed for this task. Tasks assessing extrinsic-dynamic VSA are not provided in the MaGrid app.

digital activities (Zaranis and Valla, 2017). And even very young children seem to be able to use tablets, as recently shown by Marsh et al. (2015). The authors observed that more than fifty percent of children between 0 and 5 years of age were able to drag objects on a tablet and follow shapes with their fingers on their own.

From an educational and scientific perspective, tablets seem suitable as they have been found to be effective for training and assessment of different cognitive abilities (e.g., Lowrie et al., 2014; Cornu et al., 2017). In this context, it is of particular importance that these apps consider the limited but developing, cognitive and motor skills of young children (Vatavu et al., 2015) as well as educational design principles to ensure learning (e.g., Cayton-Hodges et al., 2015). Taken together, these findings show the large potential of tablets used in education, even for young children, but also the need for the development of validated apps.

MaGrid – A Tablet-Based Early Visual-Spatial and Mathematical Training

The recently introduced tablet-based training tool MaGrid for VSA and early numerical abilities (Cornu et al., 2017; Pazouki et al., 2018) aims at meeting this challenge. MaGrid training tasks are based on established developmental models of numerical cognition (Von Aster and Shalev, 2007) as well as further findings from empirical research on visual-spatial development. Thus, they line up with the few existing digital programs for training numerical skills, which are based on generally accepted theoretical concepts and scientific evidence (e.g., "Math Garden"; Straatemeier, 2014; "Math Shelf"; Schacter et al., 2016).

MaGrid is a tablet-based app for training building blocks of early numerical abilities. It provides a wide range of training tasks (i.e., 32 number specific and simple arithmetic tasks and 16 different visual-spatial tasks). These tasks target different aspects of visual-spatial (e.g., spatial perception, (mental) rotation, spatial visualization, and visual-motor integration) and related number-specific knowledge mostly at the preschool level for children aged 4–7 years. A novelty of MaGrid is its independence of any language instructions such as text or voice-overs, which is unique so far. In addition, MaGrid combines all the advantages of computer-based training tools. It allows user-friendly easy to administer individual learning in an interactive way and provides real-time feedback. The built-in logging- and monitoring-system allows to keep track of a children's learning progress and to observe potential training-related improvements over time (Pazouki et al., 2018).

The effectiveness of MaGrid was evaluated empirically for kindergarten children (Cornu et al., 2017). In their intervention study, Cornu et al. (2017) realized a MaGrid training of VSA twice a week over a period of 10 weeks. The authors used various tasks such as *Find forms, Copy forms, Tangram, Rotation, Reproduce forms I* and *Reproduce forms II* (these tasks are used in the present study related to the 2 by 2 taxonomy of VSA), as well as *Find the pair, Figure completion, Odd-one out, Row completion, Line bisection, Figure bisection,* and *Symmetry* among others. Training effects were compared to a control group of kindergarten children (i.e., business-as-usual classroom following the Luxemburgish curriculum for kindergartens) who did not use the app. Results indicated that children who were trained with MaGrid significantly improved in some VSA (e.g., spatial orientation and visuo-motor integration) over the course of training. However, improvements in VSA were limited to the trained visual-spatial domains. The authors did not observe generalization to non-trained VSA or numerical skills. Nevertheless, this study showed the suitability of MaGrid for training VSA in kindergarten children. However, MaGrid has not yet been used as a tool for targeted assessment of VSA.

Targeted assessments are essential for the evaluation of individual abilities. However, assessments are often carried out in very artificial settings that are far from everyday life play situations. Using a tablet-based app, which has already been shown to maintain young children's interest over a longer period (Pazouki et al., 2018), may help to reduce stress in assessment situations. Thus, children's abilities may be assessed in a more playful manner (Zaranis and Valla, 2017), most probably facilitating the assessment process. Furthermore, features implemented in MaGrid, such as its language neutrality or built-in logging- and monitoring-system, may be assumed to be very promising for a fair and simplified data acquisition and monitoring of developmental processes.

In the present study and based on the above-mentioned assumptions, we modified the functionality of MaGrid so that it could be used for the assessment of VSA in kindergarten children. To this end, we chose six tasks of MaGrid, which were most closely related to the tasks Newcombe and Shipley (2015) associated with specific VSA according to their taxonomy: Two tasks each were assigned to assess intrinsic-static, intrinsic-dynamic, and extrinsic-static VSA. Please note that extrinsic-dynamic VSA (i.e., perspective taking) cannot be assessed using MaGrid because the app does not include respective tasks (see Frick et al., 2014). Therefore, we did not consider extrinsic-dynamic VSA in this study.

Using the six tasks, we evaluated whether the selected tasks conform to the taxonomy of Newcombe and Shipley (2015) applying a CFA approach (cf. Mix et al., 2018, for a similar approach). The CFA approach seems to be well suited to evaluate the structural predictions in the taxonomy of VSA by testing the fit of theoretically specified models against each other. We further investigated whether MaGrid can detect age differences in the development of VSA between three age groups of kindergarten children (youngest group: 48–58 months, intermediate group: 59–67 months, oldest group: > 68 months).

Our hypotheses were as follows: First, we expected the assignment of tasks to the categories of VSA according to the taxonomy of VSA by Newcombe and Shipley (2015) to be reflected by our empirical data as evaluated in the CFA approach. Second, provided that the visual-spatial tasks implemented in MaGrid are sensitive to reflect the hierarchical development of VSA appropriately, we further expected to observe the following specific pattern of task performance: Concerning the latent structure of the VSA according to the 2 × 2 taxonomy, we assumed to find evidence for a hierarchical development of the VSA within and across all three categories. Accordingly, older children should outperform younger children on the respective tasks within each category. Across VSA categories, task performance for intrinsic-static VSA should be better than task performance for intrinsic-dynamic VSA, which should be more pronounced than task performance in extrinsic-static VSA in all groups of children.

METHODS

Participants

Eighty-six children from four different kindergartens in the state of Baden-Wuerttemberg (Germany) participated in the study. Two children were excluded during data collection due to insufficient German language skills. Finally, data of 84 children (39 girls, mean age: M = 63.18 months, SD = 8.26 months (range 49–78 months) were included. The parents of 78 children reported that their child had German nationality. Furthermore, 56 children stated that they had experiences with tablet devices regularly.

Written informed consent was obtained from parents prior to the study besides children's verbal assent before the actual assessment. All children received a small present (e.g., a pencil and a pixie book) for their participation. The study was approved by the local ethics committee (LEK 2018/043).

Procedure

Data were collected in at least two individual testing sessions lasting ∼40 min. Testing sessions took place in a quiet and well-lit room in the respective kindergartens. Before the testing, all children were familiarized with the MaGrid app in two different ways: First, children could try out the handling of the app by playing around in the "Freeplay" mode (cf. Pazouki et al., 2018). Second, children were instructed by a tutorial video, which preceded each task and showed a visual example of solving an instance of the selected task without verbal instructions. For the assessment, we used a termination criterion to avoid repeated experiences of failure and terminated the task when a child made more than three errors in a row.

Materials

MaGrid Tasks

To assess children's VSA, we used an adapted version of MaGrid. Adaption involved several changes to the training version of the app. For example, children did not receive any feedback on their provided solutions and could only submit one solution for each item, regardless of whether they found the correct solution or not. In addition, the order of items for each task was fixed. In all tasks, items increased in task difficulty over the course of testing in order to induce variability between the tested age groups.

In the present study, we were interested in children's task performance as assessed by overall correctness in each task. To this end, an item was evaluated dichotomously as either correct or incorrect (i.e., data), resulting in a sum score for each task assessed.

Intrinsic-Static VSA

To assess children's intrinsic-static VSA, we used the tasks *Find forms* and *Close forms* of the MaGrid app (Pazouki et al., 2018). For the task *Find forms,* children were supposed to select a specific geometric form (e.g., a triangle or a rectangle), which was given in a booklet, from different distracting forms (see **Figure 1A**, in the top row) by touch-typing. This task included 16 items. In order to increase task difficulty, the number of distractors continuously increased during the task. As the number of distractors on the tablet increases, the size of the forms needed to be decreased in order to fit all forms on the display. Consequently, the size of the target form in the booklet and forms on the display varied for the more difficult trials (i.e., in 6 of 16 items). Therefore, it was explained to participating children beforehand that the size of forms in the booklet and on the tablet may differ in some trials. However, for solving the task, the shape of the form is important and not its size on the display. As *Find forms* relies basically on static pattern recognition, we assumed that the depiction of forms in different sizes should not significantly affect children's performance in intrinsic-static VSA as one would expect for active scaling processes in intrinsic-dynamic VSA.

For the task *Close forms,* a booklet was also required. The booklet showed a target form. The same form but with missing lines was displayed on the tablet in a grid. Children were asked to complete the form by drawing the missing line with their index finger (see **Figure 1A**, in the bottom row). This task also consisted of 16 items. The difficulty was increased by eliminating more lines from the given forms. In addition, the corners of a form were no longer displayed, requiring the children to create new corners to complete the forms instead of just connecting two dots in a straight line.

Intrinsic-Dynamic VSA

To assess children's intrinsic-dynamic VSA, we used the MaGrid tasks *Rotation* and *Tangram* (cf. Pazouki et al., 2018). For the *Rotation* task, children were asked to align the given form according to the orientation depicted in the booklet (see **Figure 1B**, in the top row). To this end, children were supposed to use two rotary buttons. Sixteen items were assessed. In the more difficult trials, form configuration was more specific requiring advanced visual-spatial perception.

The *Tangram* task required children to assemble various geometric forms according to a given configuration in the booklet. The forms to be assembled were presented in a random position on the tablet (see **Figure 1B**, in the bottom row). Children had to use their fingers to select a form and drag it to the correct position in relation to the other forms. Motor requirements for *Tangram* were comparably medium. *Tangram* comprises 14 items, with to-be-built configurations becoming more complex in later trials. An item was only considered to be solved correctly (and thus awarded 1 point) when all components of the form were correctly assembled (see Verdine et al., 2017, for a discussion of different coding strategies and performance on a similar spatial assembly task).

Extrinsic-Static VSA

To assess children's extrinsic-static VSA, we used the MaGrid tasks *Reproduce forms* I and II (cf. Pazouki et al., 2018). In the MaGrid task *Reproduce forms I* children had to reproduce (i.e., draw) a given geometric form in the grid of the app according to the form depicted in the booklet (see **Figure 1C**, in the top row). In sum, 25 items were assessed. The number of the given forms as well as their complexity varied between the easy and the more difficult trials.

The MaGrid task *Reproduce forms II* only differed slightly from the *Reproduce form I.* Instead of in a booklet, the target form was shown on the tablet itself in a specific position in the grid. Children were not only required to copy the given form, but they also had to reproduce the correct position in the grid (see **Figure 1C**, in the bottom row), and thus adhere to the reference frame. This task comprised 16 items. Again, more difficult tasks varied from easy tasks by using more complex forms.

The motor component for both tasks was rather high, compared to the *Tangram* task, because children had to draw on the tablet in order to copy the figure. Again, an item was only considered to be solved correctly (and awarded 1 point) when the entire form was copied correctly.

Data Analysis

Confirmatory Factor-Analysis – Structure of Early VSA

To evaluate the taxonomy of VSA suggested by Newcombe and Shipley (2015) on our data, we conducted a CFA. In the CFA, we opted on an inclusive strategy. That is, we included as many indicators per factor as possible to compensate for the relatively small sample (as recommended by e.g., Marsh et al., 1998). Items of the *Find forms* and *Close forms* tasks were considered to assess intrinsic-static VSA. Items of the tasks *Rotation* and *Tangram* were classified as assessing intrinsic-dynamic VSA. Items of the two *Reproduce forms* tasks were considered assessing extrinsic-static VSA. As all items were coded binary (i.e., correct: 1, incorrect: 0) we used the Weighted Least Squares Means and Variances (WLSMV) adjusted estimator (e.g., Li, 2016). We considered Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Tucker-Lewis Index (TLI) to evaluate model fit, with RMSEA < 0.05, CFI > 0.95, and TLI > 0.95 as cut-off criteria for a well-fitting model (Hu and Bentler, 1999). All analyses were performed in Mplus Version 8.0 (Muthén and Muthen, 2017) and SPSS (IBM®, SPSS Statistics, Version 25).

Hierarchical Development of VSA

To evaluate whether children's VSA developed hierarchically, we formed three different sub-groups according to children's age (youngest, intermediate and oldest age-group). The threshold for the oldest group was chosen because these children were old enough to enter school according to the education Act for Baden-Württemberg {Schulgesetz für Baden-Württemberg [SchG, 1983, §73 (1)]}. The second threshold was chosen to form two additional groups of similar sizes (see **Table 1**). We, therefore, assigned 27 children to the group of youngest children (i.e., 48–58 months old), 26 children were assigned to the intermediate group (i.e., 59–67 months old), and 31 children were assigned to the group of oldest children (i.e., 68–78 months old). This allowed us to investigate children's intrinsic-static, intrinsic-dynamic and extrinsic-static VSA separately for each age-group.

To test the hierarchical development of VSA in young children, we conducted both t-tests in order to investigate overall differences in children's task performance and a MANOVA evaluating the influence of age on the different categories. VSA was measured by the mean scores of correct answers for a task, with two tasks representing one ability (e.g., the intrinsic-static ability is measured by the mean score of the correct answers for *Find forms* and *Close forms*). As 56 children had prior tablet experience, we analyzed whether this experience moderated performance across tasks using the SPSS-macro PROCESS (Hayes, 2012).

The significance level was set to $p \leq 0.05$ for all analyses. Effect sizes are reported as η^2_p (medium effect $\geq$ 0.06, large effect $\geq$ 0.14, according to the recommendations of Cohen (1969, see also Richardson, 2011). Bonferroni-corrected pairwise comparisons followed-up the univariate analyses to specify significant group differences.

RESULTS

In total, data of 84 children entered the analyses. **Table 2** provides descriptive information regarding the group mean performance of the six selected MaGrid tasks. As all items were binary coded, the mean scores of the tasks indicate the percentage of correctly solved items for each task.

TABLE 1 | Sub-groups according to children's age.

Age-group	Age (months)	*M (SD)*	N	Gender (m:f)
Youngest	48–58	53.33 (2.96)	27	12:15
Intermediate	59–67	63.19 (2.67)	26	17:9
Oldest	>68	71.74 (3.47)	31	16:15

TABLE 2 | Task performance for each age group (mean correct and standard deviation).

Task	Youngest *Mean (SD)*	Intermediate *Mean (SD)*	Oldest *Mean (SD)*
Find forms	0.85 (0.15)	0.91 (0.09)	0.90 (0.08)
Close forms	0.73 (0.15)	0.80 (0.19)	0.87 (0.12)
Rotation	0.82 (0.17)	0.89 (0.15)	0.93 (0.09)
Tangram	0.36 (0.27)	0.57 (0.22)	0.71 (0.18)
Reproduce forms I	0.08 (0.14)	0.20 (0.21)	0.32 (0.27)
Reproduce forms II	0.46 (0.38)	0.76 (0.22)	0.82 (0.19)

We also looked at the correlations between tasks and found significant correlations between all tasks. **Table 3** indicated that most correlations were moderate to high (Cohen, 1988) except for the correlation between *Find forms* and *Reproduce forms I*.

Confirmatory Factor Analysis: Structure of Early VSA

We first analyzed the relative frequencies of correct and incorrect solutions in all 103 items. Items with low variance (i.e., items that were correctly or incorrectly solved by at least 90% of the children) were excluded as they did not entail sufficient information for model estimation (i.e., 44 items). Based on the remaining 59 items, we specified a three-factor model. In this model, intrinsic-static VSA were indicated by items from the *Find forms* and *Close forms* tasks (9 items in total). Intrinsic-dynamic VSA were indicated by items from the *Rotation* and *Tangram* tasks (18 items in total). Extrinsic-static VSA were reflected by items from the two *Reproduce forms* tasks (32 items in total). The model provided a good fit to the data, $\chi^2_{(1649)}$ = 1771.64, p = 0.020.02, RMSEA = 0.03, 90% CI: [0.014; 0.041], CFI = 0.98, TLI = 0.98. One additional item considered to reflect intrinsic-static VSA was dropped due to non-significant factor loading. However, model fit did not change substantially, $\chi^2_{(1592)}$ = 1717.68, p = 0.01, RMSEA = 0.03 90% CI: [0.015; 0.041], CFI = 0.98, TLI = 0.98. Taken together, these results indicate that the hypothesized three-factor structure according to Newcombe and Shipley (2015) was substantiated by the current data for kindergarten children. Item descriptions and factor loadings are presented in **Table 4**. Moreover, intrinsic-static VSA were found to be highly correlated with intrinsic-dynamic VSA (r = 0.84, p < 0.001). Similar high correlations were observed for intrinsic-static and extrinsic-static VSA (r = 0.73, p < 0.001) as well as for intrinsic-dynamic and extrinsic-static VSA (r = 0.85, p < 0.001). The final model is shown in **Figure 2**.

Hierarchical Development of Early VSA

Although not at the heart of the current research question, we first checked for overall differences in children's task performance on the three VSA. As indicated by Bonferroni-corrected t-tests, task performance was significantly better for intrinsic-static VSA (M = 0.85, SD = 0.11) than for both intrinsic-dynamic

TABLE 3 | (Pearson) correlations between MaGrid tasks.

	CF	*RO*	*T*	*RI*	*R II*
Find forms	r = 0.32*	r = 0.46*	r = 0.39*	r = 0.23*	r = 0.41*
Close forms		r = 0.44*	r = 0.63*	r = 0.48*	r = 0.58*
Rotation			r = 0.51*	r = 0.38*	r = 0.48*
Tangram				r = 0.61*	r = 0.77*
Reproduce forms I					r = 0.54*

All correlations are significant at (p < 0.05), as indicated by the asterisk (*) with CF, Close Forms; RO, Rotation; T, Tangram; RI, Reproduce Forms I; R II, Reproduce Forms II.

TABLE 4 | Descriptive statistics and factor loadings for items from the MaGrid app.

Factor	Item	% Correct	Factor loading	Factor	Item	% Correct	Factor loading	Factor	Item	% Correct	Factor loading
IS	FF12	0.488	0.765	ID	RO4	0.774	0.495	ES	RI3	0.357	0.854
IS	FF15	0.655	0.691	ID	RO5	0.679	0.821	ES	RI4	0.440	0.777
IS	FF16	0.798	0.605	ID	RO6	0.750	0.785	ES	RI5	0.429	0.869
IS	CF12	0.583	0.697	ID	RO7	0.798	0.552	ES	RI6	0.119	0.646
IS	CF13	0.405	0.824	ID	RO8	0.631	0.748	ES	RI7	0.357	0.854
IS	CF14	0.476	0.858	ID	T1	0.667	0.464	ES	RI8	0.226	0.845
IS	CF15	0.381	0.828	ID	T2	0.679	0.697	ES	RI9	0.238	0.895
IS	CF16	0.512	0.977	ID	T3	0.726	0.902	ES	RI10	0.393	0.843
				ID	T4	0.798	0.823	ES	RI11	0.179	0.716
				ID	T5	0.857	0.905	ES	RI12	0.214	0.874
				ID	T6	0.238	0.563	ES	RI13	0.143	0.701
				ID	T7	0.345	0.609	ES	RI14	0.333	0.940
				ID	T9	0.631	0.802	ES	RI15	0.274	0.959
				ID	T10	0.655	0.948	ES	RI16	0.238	0.935
				ID	T11	0.345	0.754	ES	RI17	0.214	0.898
				ID	T12	0.702	0.621	ES	RI19	0.131	0.840
				ID	T13	0.464	0.745	ES	RII1	0.690	0.775
				ID	T14	0.548	0.800	ES	RII2	0.571	0.553
								ES	RII3	0.845	0.907
								ES	RII4	0.571	0.673
								ES	RII5	0.750	0.902
								ES	RII6	0.810	0.957
								ES	RII7	0.702	0.909
								ES	RII8	0.786	0.878
								ES	RII9	0.810	0.983
								ES	RII10	0.524	0.746
								ES	RII12	0.643	0.797
								ES	RII13	0.762	0.917
								ES	RII14	0.667	0.841
								ES	RII15	0.369	0.668
								ES	RII16	0.607	0.871

IS, intrinsic-static VSA; ID, intrinsic-dynamic VSA; ES, extrinsic-static VSA.

VSA [$M = 0.73$, $SD = 0.18$, $t_{(83)} = 8.55$, $p < 0.001$] and extrinsic-static VSA [$M = 0.39$, $SD = 0.24$, $t_{(83)} = 22.01$, $p < 0.001$]. Moreover, the difference between intrinsic-dynamic and extrinsic-static VSA was also significant [$t_{(83)} = 20.10$, $p < 0.001$].

Due to the unequal distribution of boys and girls in the intermediate group, preliminary analysis by means of a MANCOVA considering sex as the covariate were conducted. There was no significant influence of the covariate sex overall [Pillai-Trace =0.031, $F_{(3,78)} = 0.820$, $p = 0.487$] as well as for the VSA categories as indicated by univariate follow-up analyses: intrinsic-static: [$F_{(1,80)} = 0.556$, $p = 0.458$; intrinsic-dynamic: $F_{(1,80)} = 0.012$, $p = 0.914$; extrinsic-static: $F_{(1,80)} = 0.807$, $p = 0.372$]. Based on these results, we are confident that the unequal distribution of boys and girls in the intermediate group did not drive our results.

To gain a better understanding of the hierarchical development of VSA, we conducted a MANOVA that indicated a significant age effect for VSA [Pillai-Trace = 0.30, $F_{(6,160)} = 4.78$, $p < 0.001$, $\eta^2_{part.} = 0.99$, see **Table 5**].

Follow-up univariate analyses indicated that there was a significant medium sized age effect for intrinsic-static VSA [$F_{(2,81)} = 5.81$, $p = 0.004$, $\eta^2_{part.} = 0.13$]. Bonferroni-corrected pairwise comparisons showed a significant difference between the youngest and oldest group only ($p = 0.003$).

For intrinsic-dynamic VSA, univariate analysis revealed a similar significant age effect with a large effect size [$F_{(2,81)} = 14.48$, $p < 0.001$, $\eta^2_{part.} = 0.26$]. Bonferroni-corrected pairwise comparisons indicated that the task performance of children in the youngest and oldest group ($p < 0.001$) differed significantly. The same applied to children in the youngest and intermediate group ($p = 0.008$).

Finally, for extrinsic-static VSA, univariate analysis indicated a significant age effect with a large effect size [$F_{(2,81)} = 14.51$, $p < 0.001$, $\eta^2_{part.} = 0.26$]. Again, Bonferroni-corrected pairwise comparisons indicated significant age differences between the youngest and intermediate group ($p = 0.003$) and the youngest and oldest group ($p < 0.001$). **Figure 3** depicts children's task performance for each category of VSA. The figure visualizes that group differences exist only between the youngest and the

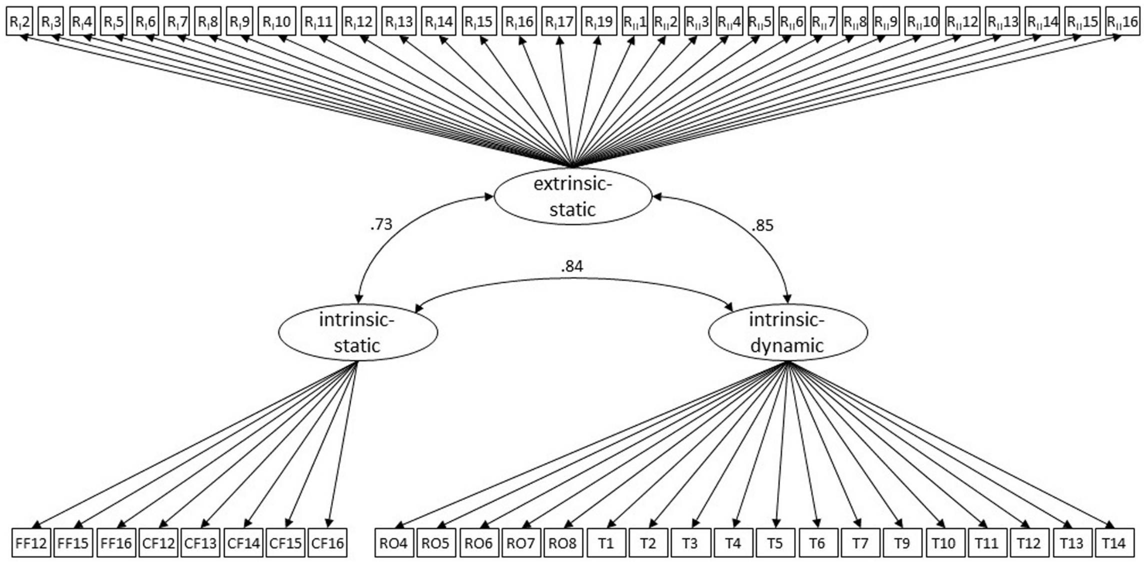


FIGURE 2 | Confirmatory factor analysis – latent structure of early VSA. The figure shows all items that were considered in the analysis. The three latent factors (i.e., intrinsic-static, intrinsic-dynamic, and extrinsic-static VSA) are derived from the 2 by 2 taxonomy of Newcombe and Shipley (2015). The figure depicts also correlations (all $p < 0.001$) among the three latent factors. To increase readability, factor-loadings and error terms of items are not displayed. Intrinsic-static VSA were measured by items of the *Find forms* and *Close forms* tasks (i.e., FF and CF), intrinsic-dynamic VSA by items of the *Rotation* and *Tangram* tasks (i.e., RO and T), and extrinsic-static VSA by items of the *Reproduce forms I* and *II* tasks (i.e., RI and RII). See text and **Table 5** for more details.

TABLE 5 | Task performance for the different age groups.

Categories	Tasks	Age group	*M*	*SD*	N	*F*	*p*	$\eta^2_{part.}$
Intrinsic-static	Find forms Close forms	Youngest	0.79	0.13	27	5.81	0.004	0.13
		Intermediate	0.86	0.11	26			
		Oldest	0.89	0.07	31			
Intrinsic-dynamic	Rotation Tangram	Youngest	0.61	0.19	27	14.48	0.000	0.26
		Intermediate	0.74	0.16	26			
		Oldest	0.82	0.11	31			
Extrinsic-static	Reproduce forms I Reproduce forms II	Youngest	0.23	0.21	27	14.51	0.000	0.26
		Intermediate	0.42	0.19	26			
		Oldest	0.52	0.21	31			

The table depicts mean correct (SD) for each age group, the number of children in each group and the test statistics for each ability.

intermediate group for intrinsic-dynamic and extrinsic-static VSA, or for the youngest and oldest group (all VSA). Crucially, no differences were observed between the intermediate and oldest group.

Results of a moderation analysis further indicated that children's prior experience with tablets did not moderate performance in intrinsic-static VSA, $\beta = -0.04$, $p = 0.137$), intrinsic-dynamic VSA ($\beta = -0.04$, $p = 0.318$), nor extrinsic-static VSA ($\beta = 0.003$, $p = 0.956$). These findings indicate that children's prior experience with tablets did not moderate the relationship between age and performance on the assessed VSA significantly.

DISCUSSION

The present study aimed at evaluating the hierarchical development of VSA from both a theoretical and a behavioral perspective. For this aim, we selected six different visual-spatial tasks of the tablet-based app MaGrid (Cornu et al., 2017; Pazouki et al., 2018): two tasks each reflecting the three categories intrinsic-static, intrinsic-dynamic, and extrinsic-static VSA of the 2 by 2 taxonomy of Newcombe and Shipley (2015).

Additionally, we adapted the functionality of MaGrid to use it for assessment purposes. Uttal et al. (2013) claimed VSA to be malleable at an early age. Therefore, accurate and reliable assessment tools are essential to both measure training success and to understand the latent structure underlying the development of VSA.

Results of the CFA indicated that the selected visual-spatial tasks reflected the respective VSA according to the taxonomy of Newcombe and Shipley (2015). Behavioral results showed that MaGrid is sensitive to detect expected age-related differences in performance between younger and older kindergarten children. In the following, we will discuss these findings in more detail

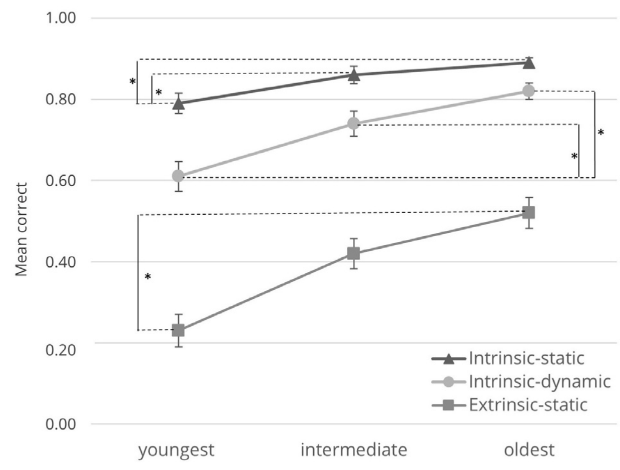


FIGURE 3 | Task performance for each age group. *M* (Mean Correct) for all three sub-groups for the tested abilities (black, intrinsic-static VSA; light-gray, intrinsic-dynamic VSA; gray, extrinsic-static VSA). Error bars reflect 1 *SE*. Significant differences with $p < 0.05$, as indicated by the asterisk (*).

beginning with the latent structure of VSA before turning to the discussion of MaGrid as an assessment tool for the development of VSA.

Latent Structure of VSA According to the 2 by 2 Taxonomy

Our CFA evaluating the structure of VSA according to the 2 by 2 taxonomy of Newcombe and Shipley (2015) indicated a good model fit for the three-factor solution reflecting the three categories of VSA of interest, this means, (i) intrinsic-static, (ii) intrinsic-dynamic, and (iii) extrinsic-static VSA. Factor loadings of all items were at an acceptable level ($\geq \sim 0.5$) alongside with a good overall fit of the model to the empirical data. CFA results suggest that the selected MaGrid tasks can be conceptualized in terms of the three (out of four) VSA as proposed by the taxonomy of Newcombe and Shipley (2015).

As regards theoretical considerations, it is important to note that we needed to exclude some items for the CFA due to insufficient variance in these items: This affected the first items of the tasks assessing the intrinsic-static (i.e., *Find forms* and *Close forms*) and the intrinsic-dynamic VSA (i.e., *Rotation* and *Tangram*). Exclusion of the first (i.e., easy) item suggests that these items may have been too easy for most children of our sample. This is in line with the observed near ceiling effects which we found for intrinsic-static VSA. Interestingly, the exclusion also affected the last items of the tasks assessing extrinsic-static VSA (i.e., *Reproduce forms I* and *Reproduce forms II).* Here, item exclusion suggests that these items may have been rather difficult for the children of our sample. Crucially, item exclusion should not negatively affect our interpretation of results. Even for the reduced number of items representing intrinsic-static VSA the statistical requirements for a just-identified factor were fulfilled, because factor loadings can be estimated independent of any particular item score (Brown, 2014).

However, analysis of response times may help to solve this issue in future studies. For instance, response times have been found to reflect specific effects of numerical processing related to visual-spatial concepts (i.e., the SNARC effect Dehaene et al., 1993). Moreover, response times and accuracy can be combined, for instance as a rate correct score (Woltz and Was, 2006), which then reflects the number of correct answers per second. It would be desirable to further pursue these avenues in future studies.

Furthermore, CFA results provided further evidence with respect to the assumptions of a hierarchical structure of the 2 by 2 taxonomy of Newcombe and Shipley (2015). CFA showed similarly high correlations between the three different factors (> 0.73). These correlations suggest that despite the division into different categories, the three VSA assessed in the current study (i.e., intrinsic-static, intrinsic-dynamic and extrinsic-static) can hardly be considered to reflect distinct constructs. Instead, they seem to represent most probably hierarchically developing VSA, and thus, help to specify the hierarchical structure of VSA, for which literature is still lacking a common definition (Eliot and Smith, 1983; Carroll, 1993; Newcombe and Shipley, 2015; Mix et al., 2016). Providing evidence of a hierarchical development and/or latent structure of VSA in the taxonomy by Newcombe and Shipley (2015) seems a major challenge for at least two main reasons: first, it may be the case that children at the age of 3 cannot solve a visual-spatial task in an assessment while they are able to solve the task during playing, in which they can master the necessary perception and action steps (Newcombe, 2018). Second, it may be that the same task requires more than one VSA to be solved (Mix et al., 2018). According to the findings of Verdine et al. (2017), spatial assembly tasks, such as the *Tangram* task are complex activities involving more than one visual-spatial ability. In the *Tangram* task, the presented form and its components need to be encoded first (i.e., requiring intrinsic-static VSA) before components need to be moved to the right position to assemble the entire form (i.e., requiring intrinsic-dynamic VSA). Both issues illustrate that theoretical assumptions of an ability and actual behavior when applying this ability do not always correspond perfectly.

MaGrid as an Age-Sensitive Assessment Tool

On the behavioral level, we observed significant age effects for all three categories (i.e., intrinsic-static, intrinsic-dynamic and extrinsic static), which was in line with our hypothesis. In all categories, we found significant differences in task performance between 4-years old (i.e., youngest group) and 6-years old (i.e., oldest group) children. Additionally, we observed significant differences between 4- and 5-years old (i.e., intermediate group) children in intrinsic-dynamic and extrinsic-static VSA. The performance of the 5- and 6-years old children did not differ significantly in any category. These results suggest MaGrid to be sensitive enough to differentiate between VSA of 4- and 6-years old children. Furthermore, the tasks assessing intrinsic-static VSA might have been too easy for children of all age groups. This might explain why only intrinsic-dynamic and extrinsic-static VSA tasks differentiated successfully between 4- and 5-years old children. However, for the latter two categories, we did not observe significant differences between the performance of 5- and 6-years old children which was contrary to our expectations. This finding might be explained by the fact that MaGrid might either not be sensitive enough to differentiate between the two age groups or the development level of the two age groups may have been too similar.

In addition to these observations, performance was higher for intrinsic-static tasks than for extrinsic-static tasks substantiating the hierarchical order of the development of these categories. This finding is particularly evident from the ceiling effects for the group of 6-years-old children for the task *Find forms*. This task requires elaborate shape recognition and abstract representation of the respective forms (i.e., intrinsic-static VSA). The task *Close forms*, which requires additional visual motor integration (Pazouki et al., 2018), demands the coordination between perceived visual input and motor output to complete the unfinished objects according to the booklet (Cornu et al., 2017).

In this context, Beery et al. (2010) observed that the development of visual motor integration was closely associated with the development of motor skills in general. In their study, they investigated this development from the ability to copy vertical lines (at the age around 2 years) and circles (at the age around 3 years) to the ability to trace horizontal lines (at the age of 3.5 years) and to connect two dots by a horizontal line (at the age of 4.5 years; Beery et al., 2010). As the youngest children in our study were 4 years and older, it is not surprising that the task *Close forms* was mastered differently well by children of different age groups.

Tasks involving intrinsic-dynamic VSA were observed to be more difficult for younger children resulting in performance differences between age groups. As dynamic VSA involve transforming and manipulating objects, such as the tasks *Tangram* and *Rotation*, they may pose higher cognitive demands. Even though it was observed in 2-year-old children that they are able to solve tasks assessing intrinsic-dynamic VSA sufficiently through perception-action skills (e.g., inserting 3D forms into appropriate slots of a box, Örnkloo and von Hofsten, 2007), this may not necessarily imply generalizability to the tasks as used in the present study (Newcombe, 2018).

Among all tasks we selected from MaGrid to assess intrinsic VSA, the *Tangram* task was the most demanding task as it requires solving visual-spatial problems by categorizing and comparing objects in relation to each other (Lin et al., 2011). Several studies indicated that tangrams inspire shape analysis, integration, and composition of objects as well as logical thinking (e.g., Olkun et al., 2005; Lin et al., 2011), and thus might be considered one of the best methods to enhance geometrical spatial thinking (Verdine et al., 2017). With its medium task difficulty and its potential involvement of other VSA (i.e., considering spatial relations of objects during visual assembly), *Tangram* seemed very suitable for assessing VSA in kindergarten children.

Finally, the most complex and difficult tasks were those assessing extrinsic-static VSA (i.e., *Reproduce forms I* and *II*), for which children of all age groups performed most poorly. The higher task demands manifested in higher variance in performance on the individual items of the tasks. Even 6-years old children in our study did not perform perfectly on these tasks and may thus not have acquired this category of VSA fully yet. This is in line with current findings showing that the understanding of spatial relations between objects and the environment as well as the size and scaling of objects improves with age and individual experiences (Newcombe and Huttenlocher, 2006; Okamoto et al., 2015).

Taken together, behavioral results indicate that basic VSA are acquired early (see Clements, 1998) and improve steadily with increasing age (Uttal et al., 2013; Newcombe et al., 2015; Cornu et al., 2017). The present results reflect that the age-related development of VSA can be measured using MaGrid. Moreover, exclusion of too easy or too difficult items (solved by almost all or no children, respectively) representing intrinsic or extrinsic VSA in the CFA only reflected the results on the behavioral level. Together, both behavioral and factor-analytical results indicated that the theoretically assumed development of VSA can be found both in the taxonomy of Newcombe and Shipley (2015) and empirically in the current data. This corroborates our theoretical understanding of the structure of VSA and their development. Although some tasks turned out to be more sensitive than others, the overall pattern of results with significant age differences for all VSA assessed corroborates the claim that kindergarten age seems central for the development of VSA (Newcombe and Frick, 2010; Cornu et al., 2017).

Limitations

When interpreting the results of the current study, some limiting aspects need to be considered. First, even though CFA models converged, our sample size is smaller than the commonly suggested lower bounds for conducting CFA of at least $N = 100$ (e.g., Anderson and Gerbing, 1988). However, there is also evidence that models can be meaningfully estimated with smaller samples. In particular, it seems that a large number of indicators per latent factor, high factor loadings, and high intercorrelations among factors may substantially decrease the required sample size (e.g., Marsh et al., 1998; Wolf et al., 2013). Given that all these aspects applied to the present data, it seems rather unlikely that sample size is a source of bias in the analyses.

Moreover, it has to be noted that several items had little to no variance and needed to be excluded from the CFA. Lack of variance was primarily caused by items that were solved correctly by almost all or no children. For future studies, it would be desirable to use additional items of medium difficulty as well as items that can differentiate also in a lower and upper ability range.

Finally, it needs to be considered that the study was cross-sectional observing VSA in children of different age levels. As such, we did not monitor the intra-individual development of children longitudinally, which means that the interpretation of developmental aspects needs to be done cautiously. Nevertheless, we think that interpretations of the development of VSA seem warranted as the present results correspond closely to previous findings (e.g., Uttal et al., 2013; Okamoto et al., 2015). Yet, future longitudinal studies would be desirable to investigate the development of (the latent structure of) VSA in more detail.

CONCLUSION

In the current study, we investigated the development and structure of VSA in kindergarten children (i.e., aged 4–6 years) using a theoretical and a behavioral approach. On the theoretical level, and based on the CFA, we found evidence to assume the latent structure of VSA as proposed in the 2 by 2 taxonomy of Newcombe and Shipley as valid (2015; but see Mix et al., 2018 for contradicting findings), and may indicated hierarchical development. On the behavioral level, we found that the development of VSA was captured by MaGrid as reflected by age differences. Moreover, we observed that the selected visual-spatial tasks fit well with the differentiation of intrinsic-static, intrinsic-dynamic, and extrinsic-static categories as proposed by this taxonomy. Thereby, these results help specify the theoretical concept of early VSA.

To conclude, the present study contributes to the literature by evaluating and validating a tablet-based assessment of early VSA. On a more theoretical level, the current study indicates that MaGrid assesses VSA on the sound theoretical basis of the taxonomy of Newcombe and Shipley (2015). On the behavioral level, the MaGrid app was found to successfully reflect individual differences in VSA in kindergarten children. In this sense, tablet-based assessments included in this educational app seem to be suitable not only for training but also for assessing VSA.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Local ethic commitee of the Leibniz-Institut für Wissensmedien (LEK 2018/043). Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

TP programmed the app for diagnostic purposes. SJ, SR, VC, CS, and KM designed the study. SJ conducted the experiment. SJ, AM, and DB analyzed the data. SJ, AM, DB, and TP wrote the original draft of the manuscript. SJ, AM, and KM reviewed and approved the final version of the manuscript. All authors contributed to the conceptualization of the study.

FUNDING

This work was funded within the framework of the Leibniz Association Pact for Research and Development and by the German Research Foundation (KL 2788/2–1). We also acknowledge support by the Open Access Publishing Fund of the University of Tübingen.

ACKNOWLEDGMENTS

We thank the participating children, their parents and institutions who made this work possible. We also thank the two Bachelor students for their assistance with data collection. We finally thank the reviewers for their insightful criticism and suggestions for improvement.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2020.00871/full#supplementary-material

DATA SHEETS S1, S2 | Raw data.

REFERENCES

Anderson, J. C., and Gerbing, D. W. (1988). Structural equation modeling in practice: a review and recommended two-step approach. *Psychol. Bull.* 103:411. doi: 10.1037/0033-2909.103.3.411

Beery, K. E., Buktenica, N. A., Beery, N. A., and Keith, E. (2010). *Developmental Test of Visual-Motor Integration*, 6th Edn. Minneapolis, StateMN: NSC Pearson.

Brown, T. A. (2014). *Confirmatory Factor Analysis For Applied Research*. New York, StateNY: Guilford Publications.

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*. Cambridge: PlaceNameplaceCambridge PlaceTypeUniversity Press.

Cayton-Hodges, G. A., Feng, G., and Pan, X. (2015). Tablet-based math assessment: what can we learn from math apps? *J. Educ. Technol. Soc.* 18, 3–20.

Chatterjee, A. (2008). The neural organization of spatial thought and language. *Semin. Speech Lang.* 29, 226–238. doi: 10.1055/s-0028-1082886

Clements, D. H. (1998). *Geometric and Spatial Thinking in Young Children*. Arlington, StateVA: National Science Foundation.

Clements, D. H., Wilson, D. C., and Sarama, J. (2004). Young children's composition of geometric figures: a learning trajectory. *Math. Think. Learn.* 6, 163–184. doi: 10.1207/s15327833mtl0602_5

Cohen, J. (1969). *Statistical Power Analysis for the Biomechanical Sciences*. New York, StateNY: L. Erbraum Associates.

Cohen, J. (1988). *Statistical Power Analysis*. Hillsdale: Erlbaum Associates.

Cornu, V., Schiltz, C., Pazouki, T., and Martin, R. (2017). Training early visuo-spatial abilities: a controlled classroom-based intervention study. *Appl. Dev. Sci.* 23, 1–21. doi: 10.1080/10888691.2016.1276835

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., and Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* 284, 970–975.

Dehaene, Stanislas, 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

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. doi: 10.1037/0012-1649.43.6.1428

Eliot, J., and Smith, placeI. M. (1983). *An International Directory of Spatial Tests*. Farmington Hills, StateMI: Cengage Learning Emea.

Franke, M., and Reinhold, S. (2007). *Didaktik der Geometrie in der Grundschule*. Amsterdem: Elsevier.

Frick, A., Möhring, W., and Newcombe, N. S. (2014). Picturing perspectives: development of perspective-taking abilities in 4-to 8-year-olds. *Front. Psychol.* 5:386. doi: 10.3389/fpsyg.2014.00386

Frick, A., and Wang, S. (2014). Mental spatial transformations in 14-and 16-month-old infants: effects of action and observational experience. *Child Dev.* 85, 278–293. doi: 10.1111/cdev.12116

Goodwin, K. (2012). *Use of Tablet Technology in the Classroom*. Sydney: NSW Department of Education and Communities.

Hayes, A. F. (2012). *PROCESS: A Versatile Computational Tool for Observed Variable Mediation, Moderation, And Conditional Process Modeling*. StateKansas: PlaceTypeUniversity of Kansas.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., and Kaufman, J. (2015). Putting education in "educational" apps: lessons from the science of learning. *Psychol. Sci. Public Interest* 16, 3–34. doi: 10.1177/1529100615569721

Hodgkiss, A., Gilligan, K. A., Tolmie, A. K., Thomas, M. S. C., and Farran, E. K. (2018). Spatial cognition and science achievement: the contribution of intrinsic and extrinsic spatial skills from 7 to 11 years. *Br. J. Educ. Psychol.* 88, 675–697. doi: 10.1111/bjep.12211

Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Model.* 6, 1–55. doi: 10.1080/10705519909540118

Hubbard, E. M., Piazza, M., Pinel, P., and Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nat. Rev. Neurosci.* 6:435. doi: 10.1038/nrn1684

Jirout, J. J., and Newcombe, N. S. (2015). Building blocks for developing spatial skills: evidence from a large, representative U.S. Sample. *Psychol. Sci.* 26, 302–310. doi: 10.1177/0956797614563338

Kozhevnikov, M., Hegarty, M., and Mayer, R. E. (2002). Revising the visualizer-verbalizer dimension: evidence for two types of visualizers. *Cogn. Instr.* 20, 47–77. doi: 10.1207/s1532690xci2001_3

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

Lee, S. A., Sovrano, V. A., and Spelke, E. S. (2012). Navigation as a source of geometric knowledge: young children's use of length, angle, distance, and direction in a reorientation task. *Cognition* 123, 144–161. doi: 10.1016/j.cognition.2011.12.015

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* 48, 936–949. doi: 10.3758/s13428-015-0619-7

Lin, C. P., Shao, Y. J., Wong, L. H., Li, Y. J., and Niramitranon, J. (2011). The impact of using synchronous collaborative virtual tangram in children's geometric. *Turkish Online J. Edu. Technology-TOJET* 10, 250–258.

Linn, M. C., and Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child Dev.* 56, 1479–1498.

Lowrie, T., Logan, T., and Ramful, A. (2017). Visuospatial training improves elementary students' mathematics performance. *Br. J. Educ. Psychol.* 87, 170–186. doi: 10.1111/bjep.12142

Lowrie, T., Logan, T., Ramful, A., and Ho, S. Y. (2014). "Do students solve graphic tasks with spatial demands differently in digital form?," *Proceedings of the Mathematics Education Research Group of Australasia*, Canberra, ACT.

Marsh, H. W., Hau, K.-T., Balla, J. R., and Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivar. Behav. Res.* 33, 181–220. doi: 10.1207/s15327906mbr3302_1

Marsh, J., Plowman, L., Yamada-Rice, D., Bishop, J. C., Lahmar, J., Scott, F., et al. (2015). *Exploring Play and Creativity in Pre-Schoolers' Use of Apps: Final Project Report*. Available online at: http://www.techandplay.org/download-report

Mix, K. S., Hambrick, D. Z., Satyam, V. R., Burgoyne, A. P., and Levine, S. C. (2018). The latent structure of spatial skill: a test of the 2$\times$ 2 typology. *Cognition* 180, 268–278. doi: 10.1016/j.cognition.2018.07.012

Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C., Hambrick, D. Z., placePing, R., et al. (2016). Separate but correlated: the latent structure of space and mathematics across development. *J. Exp. Psychol.* 145, 1206–1227. doi: 10.1037/xge0000182

Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C. J., Hambrick, D. Z., and Konstantopoulos, S. (2017). The latent structure of spatial skills and mathematics: a replication of the two-factor model. *J. Cogn. Dev.* 18, 465–492. doi: 10.1080/15248372.2017.1346658

Muthén, L. K., and Muthen, B. (2017). *Mplus User's Guide: Statistical Analysis with Latent Variables, User's Guide*. Los Angeles, StateCA: Muthén & Muthén.

Newcombe, N. S. (2018). "Three kinds of spatial cognition," *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, eds J. T. Wixted, E. A. Phelps, and L. Davachi (Hoboken, NJ: John Wiley & Sons), Vol 3, 1–31. doi: 10.1002/9781119170174.epcn315

Newcombe, N. S., and Frick, A. (2010). Early education for spatial intelligence: why, what, and how. *Mind Brain Educ.* 4, 102–111. doi: 10.1111/j.1751-228X.2010.01089.x

Newcombe, N. S., and Huttenlocher, J. (2006). "Development of spatial cognition," in *Handbook of Child Psychology*, eds W. Damon, and R. M. Lerner (Hoboken, StateNJ: Wiley).

Newcombe, N. S., Levine, S. C., and Mix, K. S. (2015). Thinking about quantity: the intertwined development of spatial and numerical cognition. *Wiley Interdiscipl. Rev.* 6, 491–505. doi: 10.1002/wcs.1369

Newcombe, N. S., and Shipley, T. F. (2015). "Thinking about spatial thinking: new typology, new assessments," in *Studying Visual and Spatial Reasoning for Design Creativity*, Ed. J. S. Gero (StateplaceBerlin: Springer), 179–192. doi: 10.1007/978-94-017-9297-4_10

Okamoto, Y., Kotsopoulos, D., McGarvey, L., and Hallowell, D. (2015). "The development of spatial reasoning in young children," in *Spatial Reasoning in the Early Years*, Ed. B. Davis (Abingdon: Routledge), 25–38.

Olkun, S., Altun, A., and Smith, G. (2005). Computers and 2D geometric learning of Turkish fourth and fifth graders. *Br. J. Educ. Technol.* 36, 317–326. doi: 10.1111/j.1467-8535.2005.00460.x

Örnkloo, H., and von Hofsten, C. (2007). Fitting objects into holes: on the development of spatial cognition skills. *Dev. Psychol.* 43:404. doi: 10.1037/0012-1649.43.2.404

Palmer, S. (1978). *Fundamental Aspects of Cognitive Representation*. Hillsdale, StateNJ: Erlbaum.

Pazouki, T., Cornu, V., Sonnleitner, P., Schiltz, C., Fischbach, A., and Martin, R. (2018). MaGrid: a language-neutral early mathematical training and learning application. *Int. J. Emerg. Technol. Learn.* 13, 4–18. doi: 10.3991/ijet.v13i08.8271

Richardson, J. T. E. (2011). Eta squared and partial eta squared as measures of effect size in educational research. *Educ. Res. Rev.* 6, 135–147. doi: 10.1016/j.edurev.2010.12.001

Sarama, J., and Clements, D. H. (2004). Building Blocks for early childhood mathematics. *Early Childh. Res. Q.* 19, 181–189. doi: 10.1016/j.ecresq.2004.01.014

Schacter, J., Shih, J., Allen, C. M., DeVaul, L., Adkins, A. B., Ito, T., et al. (2016). Math shelf: a randomized trial of a prekindergarten tablet number sense curriculum. *Early Educ. Dev.* 27, 74–88. doi: 10.1080/10409289.2015.1057462

Stiles, J., and Tada, W. L. (1996). Developmental change in children's analysis of spatial patterns. *Dev. Psychol.* 32:951. doi: 10.1037/0012-1649.32.5.951

Straatemeier, M. (2014). PlaceNameplaceMath PlaceTypeGarden: a new educational and scientific instrument. *Education* 57, 1813–1824.

Talmy, L. (2000). *Toward a Cognitive Semantics*, Vol. 2. Cambridge, StateMA: MIT press.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., et al. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychol. Bull.* 139, 352–402. doi: 10.1037/a0028446

Vatavu, R.-D., Cramariuc, G., and Schipor, D. M. (2015). Touch interaction for children aged 3 to 6 years: experimental findings and relationship to motor skills. *Int. J. Hum. Comput. Stud.* 74, 54–76. doi: 10.1016/j.ijhcs.2014.10.007

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., and Newcombe, N. (2017). Links between spatial and mathematical skills across the preschool years. *Soc. Res. Child Dev. Monogr.* 82, 1–150

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

Wolf, E. J., Harrington, K. M., Clark, S. L., and Miller, M. W. (2013). Sample size requirements for structural equation models: an evaluation of power, bias, and solution propriety. *Educ. Psychol. Measur.* 73, 913–934. doi: 10.1177/0013164413495237

Woltz, D. J., and Was, C. A. (2006). Availability of related long-term memory during and after attention focus in working memory. *Mem. Cogn.* 34, 668–684. doi: 10.3758/BF03193587

Zaranis, N., and Valla, V. (2017). "Tablet computer assisted counting and calculating activities for kindergarten children," in *Proceedings of the EDULEARN17 9th International Conference on Education and New Learning Technologies*, Barcelona, 9680–9689.

Conflict of Interest: 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.

frontiers
in Human Neuroscience

ORIGINAL RESEARCH
published: 26 May 2020
doi: 10.3389/fnhum.2020.00163

Check for updates

Early Engagement of Parietal Cortex for Subtraction Solving Predicts Longitudinal Gains in Behavioral Fluency in Children

Macarena Suárez-Pellicioni[1], Ilaria Berteletti[2] and James R. Booth[3]*

[1] Department of Educational Studies in Psychology, Research Methodology, and Counseling, The University of Alabama, Tuscaloosa, AL, United States, [2] Educational Neuroscience Program, Gallaudet University, Washington, DC, United States, [3] Department of Psychology and Human Development, Vanderbilt University, Nashville, TN, United States

There is debate in the literature regarding how single-digit arithmetic fluency is achieved over development. While the Fact-retrieval hypothesis suggests that with practice, children shift from quantity-based procedures to verbally retrieving arithmetic problems from long-term memory, the Schema-based hypothesis claims that problems are solved through quantity-based procedures and that practice leads to these procedures becoming more automatic. To test these hypotheses, a sample of 46 typically developing children underwent functional magnetic resonance imaging (fMRI) when they were 11 years old (time 1), and 2 years later (time 2). We independently defined regions of interest (ROIs) involved in verbal and quantity processing using rhyming and numerosity judgment localizer tasks, respectively. The verbal ROIs consisted of left middle/superior temporal gyri (MTG/STG) and left inferior frontal gyrus (IFG), whereas the quantity ROIs consisted of bilateral inferior/superior parietal lobules (IPL/SPL) and bilateral middle frontal gyri (MFG)/right IFG. Participants also solved a single-digit subtraction task in the scanner. We defined the extent to which children relied on verbal vs. quantity mechanisms by selecting the 100 voxels showing maximal activation at time 1 from each ROI, separately for small and large subtractions. We studied the brain mechanisms at time 1 that predicted gains in subtraction fluency and how these mechanisms changed over time with improvement. When looking at brain activation at time 1, we found that improvers showed a larger neural problem size effect in bilateral parietal cortex, whereas no effects were found in verbal regions. Results also revealed that children who showed improvement in behavioral fluency for large subtraction problems showed decreased activation over time for large subtractions in both parietal and frontal regions implicated in quantity, whereas non-improvers maintained similar levels of activation. All children, regardless of improvement, showed decreased activation over time for large subtraction problems in verbal regions. The greater parietal problem size effect at time 1 and the reduction in activation over time for the improvers in parietal and frontal regions implicated in quantity processing is consistent with the Schema-based hypothesis arguing for more automatic procedures with increasing skill. The lack of a problem size effect at time 1 and the overall decrease in verbal regions, regardless of improvement, is inconsistent with the Fact-retrieval hypothesis.

Keywords: arithmetic, subtraction, fluency, fMRI, longitudinal, children

Edited by:
Anja Ischebeck,
University of Graz, Austria

Reviewed by:
Miriam Rosenberg-Lee,
Rutgers University, Newark, United States
Catherine Thevenot,
Université de Lausanne, Switzerland
Christina Artemenko,
University of Tübingen, Germany

****Correspondence:***
Macarena Suárez-Pellicioni
mspellicioni@ua.edu

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *03 October 2019*
Accepted: *15 April 2020*
Published: *26 May 2020*

Citation:
Suárez-Pellicioni M, Berteletti I and Booth JR (2020) Early Engagement of Parietal Cortex for Subtraction Solving Predicts Longitudinal Gains in Behavioral Fluency in Children. Front. Hum. Neurosci. 14:163. doi: 10.3389/fnhum.2020.00163

INTRODUCTION

Failing math in sixth grade is a significant predictor of not graduating from high school (Belfanz et al., 2007) and math ability at age 7 predicts socioeconomic status at age 42 (Ritchie and Bates, 2013). Gaining fluency in solving single-digit arithmetic facts is an important milestone in mathematical development, freeing up working memory (Geary, 1994) and scaffolding higher-level math skills (Price et al., 2013). Despite the importance of math fluency, the neurocognitive mechanisms predicting its successful development are poorly understood.

Two hypotheses have been formulated to explain fluency development of subtraction problems. According to the Fact-retrieval hypothesis, children initially rely on slow procedures, such as counting, to solve single-digit subtractions, but with the repeated use of procedures, the problem (i.e., 5−2) and its solution (i.e., 3) are stored in long-term memory, so children shift toward retrieval (Ashcraft, 1982; Siegler, 1987). Some behavioral studies interpret the response times patterns shown by children as young as 5 years old as evidence in favor of the retrieval strategy for solving the majority of subtraction problems (Siegler, 1987). Others, relying on self-report, found that 5th graders use more retrieval and less counting to solve subtraction problems as compared to 3rd graders, who reported using more procedures (Caviola et al., 2018). According to this hypothesis, educated adults have had enough experience with arithmetic to be able to retrieve single-digit subtractions directly from memory (Siegler, 1989; Geary et al., 1993).

On the other hand, Baroody (1983) claimed that Ashcraft (1982)'s classification of retrieval as being fast and procedures as being slow was a biased assumption, and that faster response times (RTs) over development could also be explained by procedures becoming more automatic, which is the core assumption of the Schema-based hypothesis. According to Baroody (1983), children move from initial reliance on less efficient procedures such as counting to more efficient procedures including principles, heuristics or rules (e.g., $N + 0 = N$; $N \times 0 = 0$; $N - N = 0$; $N-1$ or $N + 1$ = number before or after N, respectively, in the counting sequence). Studies have suggested that procedures are solved more efficiently throughout elementary school (Woods et al., 1975) but the application of procedures seems to depend on problem type. Barrouillet et al. (2008) showed that 3rd graders reported using retrieval less frequently to solve subtractions (i.e., 19%) as compared to additions (65%; Barrouillet and Lépine, 2005) and that the retrieval of subtractions was limited to problems having a remainder of 1. Studies with adults have shown that university students retrieved only 71% (Geary et al., 1993) and 57% (Campbell and Xue, 2001) of subtractions. Procedures that adults rely on include addition reference (i.e., referring to $4 + 5 = 9$ to solve $9-4 = 5$; Peters et al., 2012; Chang et al., 2015), counting down (i.e., $9-2$ = eight, seven) and reconstruction (i.e., for $9-4$, do $10-4 = 6$; $6-1 = 5$) (Kirk and Ashcraft, 2001; Seyler et al., 2003; LeFevre et al., 2006). Studies have shown that the efficiency with which complex subtractions are solved improve even in adulthood, with older adults (i.e., 61−80 years old) being faster in applying borrowing as compared to younger adults (i.e., 18−38 years old) (Geary et al., 1993). Núñez-Peña et al. (2015) compared low and high skilled participants in a subtraction verification task in which participants reported the strategy they used to solve the problem. They found that while the two groups did not differ in the frequency of procedures vs. retrieval use, the high skilled individuals were faster and less error-prone than the less skilled ones when solving the trials for which they had reported procedural use, suggesting greater efficiency in carrying out those procedures.

Our knowledge of how subtraction problems are solved comes from behavioral studies using RTs and self-reported measures (Siegler, 1989). However, evidence has suggested that inferring mental processes from RTs can provide misleading information (Siegler, 1989) and that introspection of performance might be limited when a participant is asked to describe the strategy used when the process is fast and automatic (Ericsson and Simon, 1993; Lefevre et al., 1996). As suggested by Fayol and Thevenot (2012), participants may report using retrieval because procedures were implemented so automatically that they were not even aware of having used them. Others have claimed that the simple fact of asking about the strategies being used may alter the cognitive process, biasing participants to use those strategies that they think might be expected by the examiner (Kirk and Ashcraft, 2001). Functional magnetic resonance imaging (fMRI) can help to overcome the limitations of response times and self-reported measures by providing evidence of the underlying neurocognitive mechanisms associated with the development of subtraction fluency. Finding verbal regions of the brain to be associated with subtraction fluency gains would be compatible with the Fact-retrieval hypothesis, whereas finding quantity regions to be associated with subtraction fluency gains would be supportive of the Schema-based hypothesis. Rivera et al. (2005) found age-related increases in temporo-parietal regions, including left middle temporal gyrus (MTG) and supramarginal gyrus extending to the left intraparietal sulcus (IPS) and decreases in frontal regions such as inferior/middle frontal gyri (IFG/MFG), when 8- to 19-year-old participants solved a single-digit addition and subtraction task. Price et al. (2013) found that high school students with higher scores on a math test relied on brain regions associated with retrieval to solve single-digit additions and subtractions, whereas students with lower scores relied on brain regions associated with procedures in right IPS. Looking at the problem size effect in the brain, De Smedt et al. (2011) found that 10−12-year-old children with typical fluency relied less on quantity mechanisms in right IPS to solve small additions and subtractions, whereas children with low fluency relied on this region to solve all problems regardless of size. Polspoel et al. (2017) found that single-digit multiplications and subtractions that were reported to be solved by retrieval by 4th graders activated temporal cortex regions associated with retrieval. However, these studies have investigated brain activation by averaging across different operation types (i.e., addition and subtraction, usually). Neuroimaging evidence has shown that different operations recruit distinct neural networks (Arsalidou and Taylor, 2011; Rosenberg-Lee et al., 2011), so examining brain activation across different operations may have washed away subtraction-specific effects in the brain.

Other fMRI studies have compared subtraction processing with addition or multiplication. They have found that while additions (Rosenberg-Lee et al., 2015; Evans et al., 2016) and multiplications (Prado et al., 2011) activated verbal regions associated with retrieval, solving subtractions activated the parietal cortex, associated with procedures. Rosenberg-Lee et al. (2011) compared brain activations between single-digit addition vs. subtractions and found greater IPS activation for the latter. Prado et al. (2014) reported that children showed greater activation in the right parietal cortex when solving single-digit subtractions compared to multiplications, and this difference increased with more years of math instruction. While Prado's study can be interpreted as evidence supporting the Schema-based hypothesis, they studied maturation-related effects in the brain in a cross-sectional design that showed only a modest behavioral improvement. Concerns have been raised with the use of cross-sectional data to answer developmental questions, due to the large variability introduced by studying children from different ages, which might fail to detect or falsely suggest changes over time (Casey et al., 2005). Longitudinal studies overcome these limitations by studying the same cohort of individuals at two different time points, and constitute the recommended design (Karmiloff-Smith, 2010).

Using a longitudinal design, Artemenko et al. (2018) found that children showed reductions in frontal cortex, including MFG, from 6th to 7th grade when solving two-digit subtractions, which was accompanied by an improvement in accuracy. The reduction in frontal cortex was interpreted as less reliance on cognitive control. However, this result does not clarify whether it is fact retrieval or the use of procedures that become more efficient over time. Similar inconclusive results were found in studies showing age-related increases in both bilateral IPS and left MTG, areas associated with the use of procedures and retrieval, respectively (Rivera et al., 2005; Chang et al., 2016).

To the best of our knowledge, neuroimaging studies have not yet provided a clear picture of the underlying mechanisms responsible for fluency development in subtraction. The objective of this study was to fill this gap in the literature by answering the questions: Can reliance on verbal vs. quantity mechanisms at time 1 predict longitudinal gains in subtraction fluency, and how do these mechanisms change over time with improvement in subtraction fluency? In order to have stronger evidence for the involvement of verbal vs. quantity mechanisms, regions of interest (ROIs) were independently localized for each participant using rhyming and numerosity judgment localizer tasks, respectively. We identified ROIs implicated in the storage of phonological representations in the left MTG/STG (e.g., Prado et al., 2011), and in the access to those representations in the left IFG (e.g., Prado et al., 2011). We also localized ROIs in bilateral IPL/SPL implicated in quantity representations (e.g., Dehaene et al., 2003), and in the access of those representations in the bilateral MFG/right IFG (e.g., Arsalidou and Taylor, 2011). We then defined the extent to which children relied on verbal vs. quantity mechanisms to solve subtractions by selecting the 100 voxels showing maximal activation from each ROI, separately for small and large subtractions.

We aimed to study whether brain activation at time 1 predicts subtraction fluency gains as well as whether these neurocognitive mechanisms changed over time with fluency gains. Finding that brain activation in bilateral parietal cortex predicts the fluency gains would be compatible with both hypotheses, given that children may continue to rely on procedures that become more automatic with experience (i.e., Schema-based hypothesis), or may later shift toward retrieval (i.e., Fact-retrieval hypothesis). According to the Schema-based hypothesis, we expected to see increases in parietal cortex activation over time, suggesting that children continue to rely on procedures. However, we also expected to see decreases in bilateral MFG/right IFG over time, suggesting that procedures become more automatic (see arrow A in **Figure 5**; Schema-based). According to the Fact-retrieval hypothesis, we expected to see decreases in parietal cortex and increases in temporal cortex over time. It is possible that this process is accompanied by increases in left IFG activation over time, given that the implementation of retrieval strategy might be effortful in its early stages (Geary et al., 1996a; i.e., see arrow B in **Figure 5**; Fact-retrieval). Finally, there is a third possibility. Considering evidence suggesting that by age 10 retrieval may be the dominant strategy to solve single-digit arithmetic problems (Ashcraft and Fierman, 1982), it might be the case that children have already shifted toward retrieval at time 1, in which case we expect to see activation in temporal cortex early on to predict fluency gains. In this scenario, we expect children to show increases in temporal cortex activation over time, suggesting that they build their storage of subtraction facts in long-term memory. This might be accompanied by decreases in left IFG over time, suggesting that the retrieval becomes less effortful as the representations become more robust (Prado et al., 2014; see arrow C in **Figure 5**; Fact retrieval).

MATERIALS AND METHODS

Participants

Whole Sample

Sixty-five 3rd to 8th graders were recruited from schools in the Chicago metropolitan area to participate in the study. This dataset has been deposited in OpenNeuro (10.18112/openneuro.ds001486.v1.1.0) and a detailed description of the dataset is provided in Suárez-Pellicioni et al. (2019b). Timepoint 1 of this dataset is the basis of other publications by our research group, including (Berteletti et al., 2014; Demir-Lira et al., 2014, 2015; Prado et al., 2014; Berteletti and Booth, 2015a,b; Demir-Lira et al., 2016). The longitudinal data of this dataset is the basis of other publications including Suárez-Pellicioni and Booth (2018), Suárez-Pellicioni et al. (2018), Suárez-Pellicioni et al. (2019a). None of them have looked at longitudinal gains in subtraction fluency, which constitutes the objective of this study.

All participants were native English speakers, right-handed, were free of past and present psychiatric disorders including Attention Deficit Hyperactivity Disorder (ADHD), neurological disease or epilepsy. According to parental report, no participant had hearing impairments, uncorrected visual impairment, was born prematurely (less than 36 weeks), was taking medication affecting the central nervous system or had any contraindication for being scanned, such as having braces. Participants had no history of intellectual deficits, all of them scoring above 85 standard score (hereinafter, SS) on the Full IQ scale of the Wechsler Abbreviated Scale of Intelligence – WASI (Weschler, 1999). All participants scored above 71 SS on the Math Fluency subtest from the Woodcock-Johnson III Test of Achievement (WJ-III; Woodcock et al., 2001) and above 85 SS on the average of Word Attack and Word Identification tests of the WJ-III. Children and their parents or guardians provided written consent to participate in the study. Parents were compensated $20 per hour for their time. All experimental procedures were approved by the Institutional Review Board at Northwestern University.

Data from six participants had to be excluded because of having excessive movement in the scanner. Excessive movement was defined as more than 10% of the total volumes replaced or more than five consecutive volumes replaced in a given run (for more details, see Section "fMRI Data Analysis"). There was no correlation between number of volumes replaced and age ($r = -0.12$, $p = 0.43$) or improvement ($r = 0.09$, $p = 0.51$).

Data from another six participants were excluded for showing accuracy below 50% in the small condition of the subtraction task solved inside the scanner either at time 1 or time 2 (for more specific information see section "Subtraction Task Behavioral Results"). Six additional participants had to be excluded for showing accuracy below 33% for the control condition (i.e., blue square). One participant was excluded for being left-handed.

The final sample consisted of 46 participants[1] who were tested longitudinally, with sessions being approximately 2 years apart. More detailed information about the sample is given in **Table 1**.

Improvement Groups

Two groups were created based on improvement on the subtraction task solved inside the scanner: improvers and non-improvers (see section "Experimental Task: Single Digit Subtraction" for a description of the subtraction task and its conditions). To form the groups, we first calculated the difference in means of response times between time points (i.e., Time 2-Time 1) for large subtractions. In order to account for initial differences in performance, we regressed time 1's response times out from the difference score, saving the residuals. These residuals represented the difference in response times after initial levels have been accounted for. Then, we created two groups based on the median-split of these residuals: improvers (n = 23) and non-improvers (n = 23). The decision of using large subtractions was made given the simplicity of the small subtractions in our study, with half of the problems having a remainder of 1 (e.g., 3 – 2 = 1), the largest remainder being 3 (e.g., 5 – 2 = 3), and that 40% of the problems included minuends smaller or equal 5. More detailed information about these two groups is shown in **Table 1**. The two groups did not differ in age at time 1, age at time 2, time between sessions, sex distribution, reading skill, verbal WM, visuo-spatial WM, verbal IQ, or performance IQ (all p-Values above.22; all measured using age-adjusted norms). For more information on differences in performance between these groups see section "Improvement Groups' Performance" and **Figure 6**.

[1]The following participants were included in this study: 2, 5, 6, 7, 8, 9, 10, 11, 12, 13, 16, 20, 22, 27, 34, 35, 36, 40, 44, 45, 48, 49, 50, 53, 54, 56, 57, 59, 60, 65, 67, 69, 70, 71, 75, 76, 77, 83, 86, 89, 90, 93, 95, 96, 103, and 106.

Standardized Measures

Reading skill was measured as the average of standard scores on the Word Attack and the Word Identification subtest from the Woodcock-Johnson III Test of Achievement (WJ-III; Woodcock et al., 2001) at time 1. The Word Attack requires oral reading of pseudo-words, while the Word Identification test requires oral reading of isolated letters and real words.

Verbal working memory (WM) was measured by the Listening Recall subtests of the Automated Working Memory Assessment (AWMA; Alloway et al., 2007). This subtest requires children to decide whether a sentence is true or false and also to remember the final word of the sentence. Thus, children are asked to store the final word of the sentence, as they process an increasing number of new sentences. The item is scored as correct if children recall the correct word or words in the correct order.

Visuo-spatial WM was measured with the Spatial Recall subtest of the AWMA (Alloway et al., 2007). In this test, children view pictures of two shapes where the shape on the right has a red dot near it and they need to identify whether the shape on the right is the same as the shape on the left when rotated in two dimensions, or whether it is the mirror image. At the end of the trial, individuals are asked to remember the position of the red dot and to answer by pointing to a picture with three possible positions marked. The number of shape pairs to be compared increases as children proceed through the test, and participants must recall the correct position of all the red dots in the correct temporal order.

Intelligence was measured using both the Verbal and Performance subscales of the Wechsler Abbreviated Scale of Intelligence (WASI; Weschler, 1999). Verbal IQ was measured with the Vocabulary subtest, in which the participants have to define words, and with the Similarities subtest, in which the participants are presented with two words that represent common objects or concepts and they have to describe how they are similar. Performance IQ was measured with Block Design and Matrix Reasoning subtests of the WASI. The Block Design requires the participants to use red-and-white blocks to re-create, within a specified time limit, a model design. In the Matrix Reasoning subtest, participants view an incomplete series or matrix and select the response option that completes it logically.

Scanner Tasks

Rhyming Judgment Localizer Task to Identify Verbal Regions in the Brain

In the rhyming judgment task, two written monosyllabic English words were sequentially presented and participants had to decide whether the words rhymed or not. To ensure that participants relied on phonology to solve the task, and not orthography, we created four conditions in which pairs of words had: (1) similar orthography and similar phonology (i.e., O + P +;

TABLE 1 | Demographic characteristics and standardized scores.

	Whole sample (*n* = 46)	Improvers (*n* = 23)	Non-improvers (*n* = 23)	Group differences
Age at T1 session (years)	11.2 (1.5)	11.1 (1.6)	11.3 (1.5)	$t(44) = -0.56$, $p = 0.58$
Age at T2 session (years)	13.4 (1.6)	13.3 (1.8)	13.6 (1.6)	$t(44) = -0.60$, $p = 0.55$
Time between sessions (years)	2.1 (0.2)	2.1 (0.2)	2.2 (0.2)	$t(44) = -0.63$, $p = 0.54$
Female/male ratio	25/21	15/8	10/13	$X^2 = 2.20$, $p = 0.14$
Reading at T1 (SS)	107.0 (10.3)	107.7 (10.6)	106.3 (10.3)	$t(44) = 0.45$, $p = 0.65$
Verbal WM at T1 (SS)	103.0 (13.4)	103.8 (15.6)	102.3 (11.1)	$t(44) = 0.39$, $p = 0.70$
Visuo-spatial WM at T1 (SS)	106.0 (13.0)	106.9 (15.0)	104.9 (10.8)	$t(44) = 0.52$, $p = 0.60$
Verbal IQ at T1 (SS)	114.0 (16.0)	115.7 (16.3)	111.3 (15.3)	$t(44) = 0.94$, $p = 0.35$
Performance IQ at T1 (SS)	110.2 (15.2)	113.0 (16.1)	107.5 (14.1)	$t(44) = 1.23$, $p = 0.22$

Participant's age at each time point, time between sessions, number of females and standard scores (i.e., adjusted for age norms) in reading skill, working memory (WM) and intelligence (IQ) at time 1 for the whole sample (n = 46), for the improvers (n = 23), and non-improvers (n = 23). The last column indicates statistical values for the comparison between improvers and non-improvers. SS, Standard score; T1, Time 1; T2, Time 2. See section "Standardized Measures" for a description of the tests used to measures reading skill, verbal WM, visuo-spatial WM, verbal IQ, and performance IQ.

e.g., *dime–lime*; 12 trials); (2) similar orthography but different phonology (i.e., O + P-; e.g., *pint–mint;* 10 trials); (3) different orthography but similar phonology (i.e., O−P +; e.g., *jazz–has;* 10 trials); (4) different orthography and different phonology (i.e., O−P−; e.g., *press–list;* 14 trials). The O + P + and O−P− constituted the non-conflicting conditions, given that orthographic information was consistent with the right answer, whereas the O-P + and O + P− conditions constituted the conflicting conditions because orthographic information was inconsistent with the right answer. **Figure 1A** shows an example of an O + P- condition of the rhyming judgment task. The control condition consisted of a blue square that was presented for the same duration as the experimental conditions and children were asked to press a button when the square turned red (**Figure 1E**). Stimuli were presented in a single run, lasting approximately 7 min. All participants received trials in the same order.

Numerosity Judgment Localizer Task to Identify Quantity Regions in the Brain

Participants were sequentially presented with two dot arrays and their task was to decide which of them had more dots. The task comprised 24 easy (i.e., compare 12 vs. 36 dots), 24 medium (i.e., 18 vs. 36), and 24 hard (i.e., 24 vs. 36 dots) trials. The first dot array was composed of the larger number of dots in half of the trials, while it was composed of the smaller number of dots in the other half. To ensure that participants' judgments were based on differences in quantity rather than cumulative surface area, the distribution of dot sizes was biased toward smaller dots in large arrays and bigger dots in small arrays. However, totally equating the cumulative surface area between small and large arrays by entirely biasing the distribution of single dot sizes (100% bias) may have led participants to use single dot sizes as a cue for their judgments. Therefore, we found a trade-off (50% bias) between equating as much as possible the cumulative surface areas and the distributions of single dot sizes in each pair. **Figure 1B** shows an example of an easy condition of the numerosity judgment task. The control condition consisted of a blue square that was presented for the same duration as the experimental conditions and children were asked to press a button when the square turned red (**Figure 1E**). Stimuli were divided into two runs, lasting approximately 4 min each. All participants received trials in the same order within each run.

Experimental Task: Single-Digit Subtraction

Participants were presented with a single-digit subtraction problem followed by a proposed solution and were asked to decide whether the proposed solution was true or false. Problems were broken down into small (**Figure 1C**) and large (**Figure 1D**) single-digit problems. Small subtractions (12 problems) were characterized by having a small difference (i.e., 1, 2, or 3) between the first and second term of the subtraction (e.g., 3 - 2), regardless of the first term size. In large subtractions (12 problems), the first term was relatively large (i.e., 6, 7, 8, or 9), as was the difference between the first and second terms (i.e., 3, 4, 5, or 6; e.g., 9 - 4). Each problem was repeated twice with a true solution and once with a false solution, yielding a total of 72 trials. False solutions were constructed by adding 1 or 2 to the correct solution (e.g., 8 − 2 = 7), or by subtracting 1 from the correct solution (e.g., 8 − 5 = 2). Problems involving 0 (e.g., 3 − 3; 3 − 0) or 1 as the second operand (e.g., 3 - 1) and ties (e.g., 6 - 3) were only used in the practice session. The control condition consisted of a blue square that was presented for the same duration as the experimental conditions and children were asked to press a button when the square turned red (**Figure 1E**). Stimuli were divided into two runs, lasting approximately 4 min each. All participants received trials in the same order within each run.

Experimental Protocol

First, informed consent was obtained from the children and their parents or guardians, and then standardized tests were administered. Children then had a practice session in which they practiced all trial types and learned to minimize head movement in a mock fMRI scanner. For the rhyming and numerosity localizer tasks, the practice session consisted of twelve trials of each condition. For the subtraction task, twenty-four problems with a correct proposed solution and 24 problems with a false proposed solution were included in the practice session. For all the tasks, the items used for the practice session were different from the ones used for the scanning session.

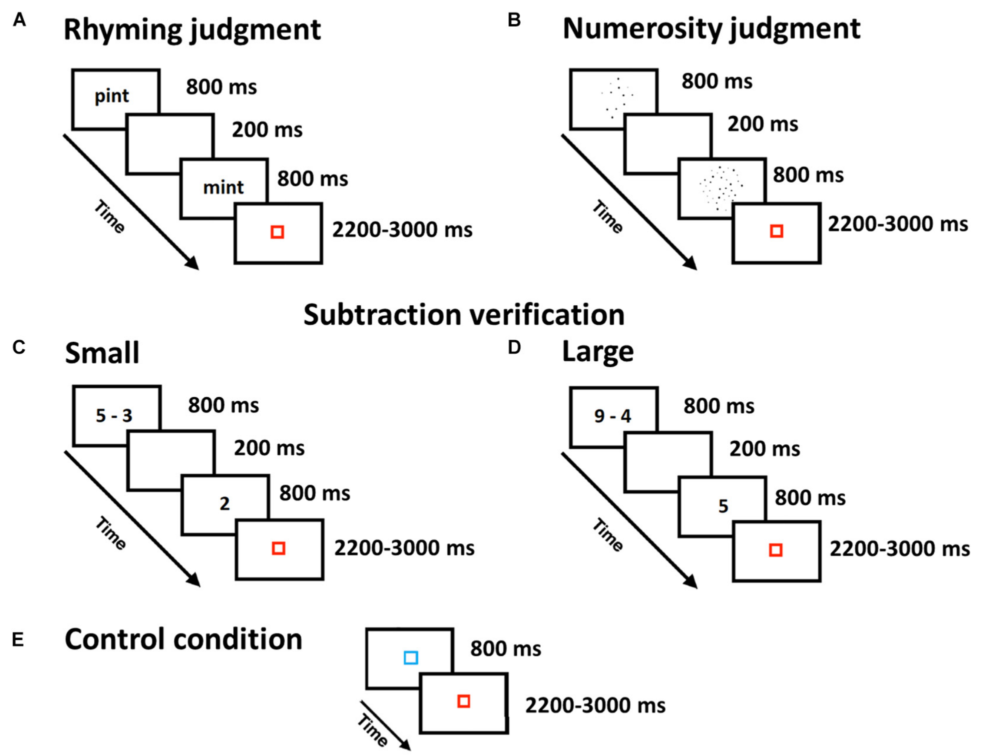


FIGURE 1 | Experimental tasks and their timing. Localizer task: **(A)** The rhyming judgment task was used to identify verbal regions of the brain in which participants had to respond whether pairs of words rhymed or not. **(B)** The numerosity judgment task was used to identify quantity regions of the brain in which participants had to indicate which of the two sets of dots had a greater number. Subtraction task: Single-digit verification task, including **(C)** small and **(D)** large subtractions. **(E)** Control condition common to all tasks, in which participants had to press a button when the blue square turned red.

The actual scanning session took place within a week of the practice session. In the fMRI scanner, participants performed one run of the rhyming judgment task, two runs of the numerosity judgment task and two runs of the subtraction verification task. The order of the tasks and the runs was counterbalanced across participants. The timing and order of trial presentation were optimized for estimation efficiency using optseq2[2]. Behavioral responses were recorded using an MR-compatible keypad and participants responded with their right hand. Participants responded with their index finger if the two words rhymed, if the first array of dots had more dots, if the proposed solution for the subtraction problem was correct, or when the blue square from the control condition turned red. Participants used their middle finger if the two words did not rhyme, if the second array of dots had more dots, or if the proposed solution for the subtraction problem was incorrect. Stimuli were generated using E-prime software (Psychology Software Tools, Pittsburgh, PA, United States) and projected onto a screen that was viewed by the participants through a mirror attached to the head-coil.

Stimulus Timing

Stimulus timing was identical for all tasks. A trial started with the presentation of a first stimulus (i.e., first word, first array of dots, or subtraction operation) for 800 ms followed by a blank screen for 200 ms. A second stimulus (i.e., second word, second array of dots, or proposed solution for the subtraction operation) was presented for 800 ms, and followed by a red fixation square for 200 ms. Variable periods of fixation, ranging from 2200 to 3000, were added after each trial in order to help with convolution, during which a red square was presented. Participants could respond as soon as the second word was presented until the beginning of the next trial. As for the control condition, the blue square was presented for 800 ms followed by a red fixation square lasting 2200-3000 ms. The run ended with 22 s of passive visual fixation in order to aid in deconvolution of the final trials.

fMRI Data Acquisition

Images were collected using a Siemens 3T TIM Trio MRI scanner (Siemens Healthcare, Erlangen, Germany) at Northwestern University's Center for Advanced MRI. The fMRI blood oxygenation level dependent (BOLD) signal was measured with a susceptibility weighted single-shot echo planar imaging (EPI) sequence. The following parameters were used: TE = 20 ms, flip angle = 80°, voxel size: 1.7 × 1.7 × 3 mm, matrix size = 128 × 120 × 37, field of view = 220 × 206.25 × 111 mm, slice thickness = 3 mm (0.48 mm gap), number of slices = 32, TR = 2000 ms. Before functional image acquisition, a high resolution T1 weighted 3D structural image was acquired for

[2]http://surfer.nmr.mgh.harvard.edu/optseq/

each subject, with the following parameters: TR = 2300 ms, TE = 3.36 ms, matrix size = 256 × 256, field of view = 240 mm, slice thickness = 1 mm, number of slices = 160.

fMRI Data Analysis

Preprocessing

Data analysis was performed using SPM8[3]. The first six images of the run were discarded to allow for T1 equilibration effects. The remaining functional images were corrected for slice acquisition delays, realigned to the first image of the run to correct for head movements, and spatially smoothed with a Gaussian filter equal to twice the voxel size (4 × 4 × 8 mm^3 full width at half maximum). Prior to normalizing images, we used ArtRepair ([4]Mazaika et al., 2009) to identify outlier volumes with more than 1.5 mm in volume-to-volume movement in any direction, or with more than 4% deviation from the mean global signal. The outlier volumes were repaired by interpolation between the nearest non-outlier volumes. All participants had less than 10% of the total number of volumes replaced and less than 5 volumes replaced in a row. Interpolated volumes were then partially de-weighted when first-level models were calculated on the repaired images (Mazaika et al., 2007). Functional volumes were co-registered with the segmented anatomical image and normalized to the standard T1 Montreal Neurological Institute (MNI) template volume (normalized voxel size, 2 × 2 × 4 mm^3).

fMRI Processing

Event-related statistical analysis was performed according to the general linear model. Activation was modeled as epochs with onsets time-locked to the presentation of the first stimulus in each trial. All epochs were convolved with a canonical hemodynamic response function. The time series data were high-pass filtered (1/128 Hz), and serial correlations were corrected using an autoregressive AR model. Considering that improvement groups did not significantly differ in accuracy at either time point (see section "Improvement Groups' Performance" for more details) and in order to equate for power in the analysis, all children's responses (i.e., correct and incorrect) were included in the model.

Regions of Interest Definition

Regions of interests were defined base on a sample of 40 participants. Six participants[5] had to be excluded for ROI definition because of having low accuracy in the rhyming judgment task (n = 1) and due to excessive movement in both localizer tasks (n = 5). Five combined ROIs were created, combining functional and anatomical ROIs. These combined ROIs were created by identifying the regions showing activation for the rhyming and numerosity judgment localizer tasks within fronto-temporal and fronto-parietal anatomical regions, respectively. The rationale for using combined ROIs, instead of only anatomical ones, was to be more confident of the underlying cognitive mechanisms (i.e., verbal vs. quantity) engaged during subtraction solving.

To localize quantity regions in the brain we identified, for each participant, the voxels that showed greater activation for all dot pairs of the numerosity judgment task as compared to the control condition, at time 1. In a second-level analysis, these individual contrasts were submitted to a one-sample t-test across all participants. Given extensive evidence suggesting that the bilateral intraparietal sulci (IPS) is the crucial neural substrate for numerical magnitude processing (Pinel et al., 2001; Dehaene et al., 2003; Sokolowski et al., 2017), we used the bilateral IPL/SPL anatomical regions to ensure coverage of the IPS. We then constrained the brain activation elicited by the numerosity judgment localizer task within the anatomical bilateral IPL/SPL and took this combined ROI as the region responsible for quantity representations in left (**Figure 2A**) and right (**Figure 2B**) parietal cortices. All anatomical regions were defined using the anatomical automatic labeling (aal) template, which is part of the WFU pickatlas tool (Maldjian et al., 2003). Given previous evidence suggesting that the left IPL/SPL plays a crucial role in calculation (Simon et al., 2002; Rivera et al., 2005; Price et al., 2016), we considered left and right IPL/SPL as separate ROIs, in order to explore hemispheric differences.

To localize verbal regions in the brain we identified, for each participant, the voxels that showed greater activation for all word pairs of the rhyming judgment task as compared to the control condition, at time 1. In a second-level analysis, these individual contrasts were submitted to a one-sample t-test across all participants. Based on extensive literature suggesting that left lateral temporal cortex is implicated in housing phonological representations (Booth et al., 2002, 2003, 2004; Prado et al., 2011, 2014), we constrained the brain activation elicited by this contrast within the anatomical left middle and superior temporal gyri (MTG/STG) and considered this combined ROI to represent the storage of verbal representations (**Figure 2D**).

While different anatomical regions were used to identify the storage of verbal vs. quantity representations, the previous literature on the brain regions involved in *accessing* those representations, especially quantity representations, is less robust. For this reason, we decided to use the same anatomical region, the bilateral frontal cortex (i.e., inferior, middle and superior frontal gyri), to identify the regions involved in accessing verbal and quantity representations. When comparing the brain activation of all dot pairs of the quantity task vs. the control condition within the bilateral frontal cortex, we found three clusters that reached significance: one in the left middle frontal gyrus (MFG; **Figure 2C**, left), one in right IFG (**Figure 2C**, middle), and one in right MFG (**Figure 2C**, right), which were taken as the ROIs involved in accessing quantity representations. This goes in line with Arsalidou's meta-analyses suggesting that these regions are active for calculation in adults (Arsalidou and Taylor, 2011) and in children (Arsalidou et al., 2018), and for non-symbolic quantity processing (Sokolowski et al., 2017). These regions have also been found to be more active for subtraction as compared to additions (De Smedt et al., 2011; Rosenberg-Lee et al., 2015), for subtractions as compared to a control condition (Kawashima et al., 2004; Evans et al., 2016) and for arithmetic problems reported to be solved by procedures as compared to those

[3]www.fil.ion.ucl.ac.uk/spm
[4]https://cibsr.stanford.edu/tools/human-brain-project/artrepair-software.html
[5]Participants 50, 56, 65, 71, 83, and 96 were excluded from the ROI definition.

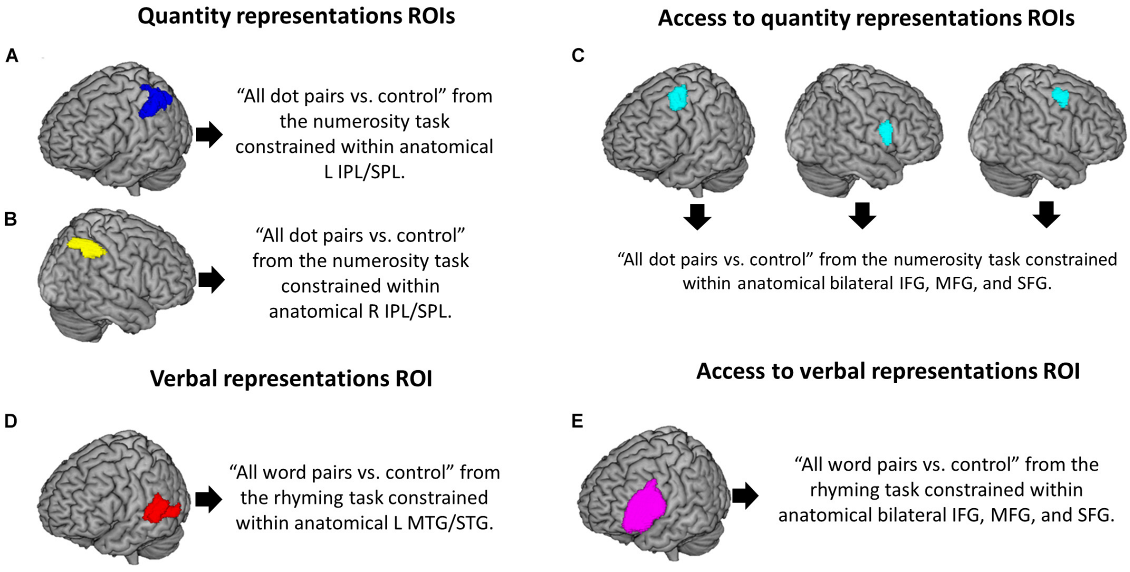


FIGURE 2 | Visualization of regions of interest. Regions of the brain implicated in: quantity representations in **(A)** left and **(B)** right IPL/SPL; access to quantity representations in **(C)** bilateral MFG/right IFG (including left MFG, right IFG, and right MFG); **(D)** storage of verbal representations in left MTG/STG; and **(E)** access to verbal representations in left IFG.

reported to be retrieved (Grabner et al., 2009; Polspoel et al., 2017). Given that the role of these three regions in arithmetic processing is not yet clear and that we did not have specific predictions for each area, we treated the three clusters as a single ROI (hereinafter, bilateral MFG/right IFG). To the best of our knowledge, this is the first study that identifies frontal regions involved in quantity processing by means of a localizer task and uses brain activation from these regions to predict subtraction fluency gains.

When constraining the brain activation of all word pairs of the rhyming judgment task vs. the control condition within the bilateral frontal cortex, we found that a cluster in the left inferior frontal gyrus (IFG) was the only one that reached significance (**Figure 2E**). This finding goes in line with extensive previous evidence suggesting that left IFG is responsible for accessing verbal representations (Poldrack et al., 1999; Rickard et al., 2000; Bookheimer, 2002; Booth et al., 2003, 2004; Prado et al., 2011, 2014; Fedorenko et al., 2012; Andin et al., 2015; Pollack and Ashby, 2017). As shown in **Figure 2**, the ROIs involved in accessing quantity (2C) and verbal (2E) representations showed no overlap. More information about these combined ROIs is given in **Table 2**.

Statistical significance for creating these combined ROIs was defined using Monte Carlo simulations in AFNI's 3dClustSim program (December, 2015[6]; with SPM's data smoothness parameters, autocorrelation function [ACF] = 0.45, 4.14, 11.02). 3dClustSim carries out a user-specified number of Monte Carlo simulations of random noise activations at a particular voxel-wise alpha level within a masked brain volume. Following the suggestions made by Eklund et al. (2016) regarding the inflated statistical significance achieved using some packages (i.e., SPM, FSL, and AFNI), we used 3dClustSim's most recent version (December, 2015). We used 3dFWHMx to calculate the smoothness of the data for every participant, using a spatial ACF, and then averaged those smoothness values across all participants. This averaged smoothness value was then entered into 3dClustSim to calculate the cluster size needed for significance for a given anatomical mask. Cluster sizes of 92, 53, and 53 were needed to reach significance for the bilateral frontal cortex, left MTG/STG and bilateral IPL/SPL anatomical regions, respectively. Clusters exceeding these size thresholds, at a cluster-wise threshold of $p = 0.05$ and voxel-wise threshold of $p = 0.005$, were deemed significant.

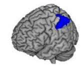
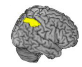
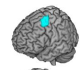
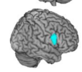
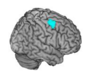
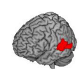
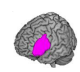
ROI Analysis

The 100 voxels showing maximal activation for the contrast "small subtractions vs. control" and "large subtractions vs. control" at time 1 were extracted for every participant from each of the five ROIs described above (i.e., left IPL/SPL, right IPL/SPL, bilateral MFG/right IFG, left MTG/STG, and left IFG).[7] The selection of brain voxels showing maximum activation at the individual level has been suggested to provide higher sensitivity and selectivity, being better able to detect effects and distinguish between conditions (Fedorenko et al.,

[6]http://afni.nimh.nih.gov/

[7]Note that the variable of interest in this study is the extent to which participants relied on quantity representation mechanisms, indicating the use of calculation-based strategies, so we extracted brain activation *during the subtraction task* (i.e., top 100 voxels) from ROIs in parietal cortex that were identified during the numerosity judgment task using the contrast "all dots vs. control". In contrast, in Suárez-Pellicioni and Booth (2018), we explored the role of quantity representation at time 1 in predicting math fluency gains, with parietal activation *during the Numerosity judgment task* (i.e., contrast "hard vs. easy") being our covariate of interest.

TABLE 2 | Information for regions of interest.

Localizer contrast	Anatomical constraint	K	aal	~BA	MNI coordinate			*Z*-value	Cluster in Figure 2
					X	*Y*	*Z*		
Dot pairs > control	Left IPL/SPL	580	Left IPL/SPL	7/40	−34	−37	38	5.0	
					−40	−41	42	5.0	
					−42	−27	42	4.9	
	Right IPL/SPL	286	Right IPL/SPL	7/40	46	−37	54	5.5	
					24	−63	50	5.3	
					30	−53	46	5.1	
	BilateralIFG/MFG/SFG	202	Left MFG	6	−30	−7	66	5.9	
					−28	−1	58	4.5	
					−26	7	62	3.7	
		130	Right IFG oper	44	54	7	26	5.6	
					58	11	18	5.1	
		133	Right MFG	6	32	−1	58	5.5	
Word pairs > control	Left MTG/STG	495	Left MTG/STG	21/22	−56	−35	2	5.3	
					−42	−61	−2	5.0	
					−50	−67	−2	4.4	
	Bilateral IFG/MFG/SFG	1546	Left IFG	44/45/47	−48	13	26	7.4	
					−38	29	6	6.6	
					−46	27	18	6.3	

*Localizer contrast and anatomical constraint used to create the combined ROIs. Detailed information of the combined ROIs including cluster size (k), corresponding region based on anatomical automatic labeling (aal), approximate Brodmann areas (~BA), MNI coordinates of the peaks, Z-values, and corresponding cluster in **Figure 2**. MTG/STG, middle and superior temporal gyri; IFG, inferior frontal gyrus; IPL/SPL, inferior and superior parietal lobules; MFG, middle frontal gyrus; SFG, superior frontal gyrus.*

2010; Nieto-Castañón and Fedorenko, 2012), as compared to traditional group-based analyses that tend to overestimate overlap across participants and underestimate functional specificity (Fedorenko and Kanwisher, 2009). **Figure 3** shows the cluster overlap in the five ROIs across participants at time 1, separately for small and large subtractions. Brain activation at time 2 was also extracted from these clusters identified at time 1 in order to study changes in brain activation over time.

Parameter estimates (or β weights) associated with the two contrasts were extracted at the individual level using MarsBars. Subsequently, the extracted data were submitted to SPSS 22 (IBM, SPSS Statistics, IBM Corporation, NY, United States) for statistical testing.

Statistical Analyses on Brain Activations During Subtraction Task Solving

Brain activations elicited at time 1 while solving small and large subtractions were separately extracted from the five ROIs (i.e., left IPL/SPL, right IPL/SPL, bilateral MFG/right IFG, left MTG/STG, and left IFG), resulting in 10 variables (i.e., neural problem size effect).

In analysis 1, we studied the role of brain activation at time 1 while solving small and large subtraction problems in predicting math fluency gains. To this aim, we ran a mixed ANOVA including Improvement groups (i.e., improvers; non-improvers) as the between-subjects factor and Problem size (i.e., small, large) × ROI (i.e., L IPL/SPL, R IPL/SPL, bilateral MFG/right IFG, left MTG/STG, and left IFG) at time 1 (i.e., the neural problem size effect) as the within-subjects factors. Participants' age at time 1 and large subtractions' accuracy at time 1 were included as covariates. **Figure 4A** shows an illustration of the between-subjects factor, within-subjects factors, and covariates included in this analysis.

In analysis 2, we explored the changes over time in brain activation associated with subtraction fluency improvement. We ran a mixed ANOVA including Improvement groups (i.e., improvers; non-improvers) as the between-subjects factor and Problem size (i.e., small, large) × ROI (i.e., L IPL/SPL, R IPL/SPL, bilateral MFG/right IFG, left MTG/STG, and left IFG) × Time (time 1, time 2) as the within-subjects factors. Participants' age at time 1 and large subtractions' accuracy at time 1 were included as covariates. **Figure 4B** shows an illustration of the between-subjects factor, within-subjects factors, and covariates included in this analysis.

Figure 5 shows an illustration of the findings supporting each hypothesis tested in this study, expected at time 1 and expected for the changes in brain activation (time 2 vs. time 1). Finding that brain activation in bilateral IPL/SPL predicts fluency gains would be compatible with both hypotheses given that children may initially rely on parietal-based procedures and continue to do so over time. However, these procedures may become more automatic (i.e., Schema-based hypothesis), or children may initially rely on procedures but later shift toward retrieval (i.e., Fact-retrieval hypothesis). In the first case, illustrated in arrow A in **Figure 5**, we expected to see increases in bilateral IPL/SPL activation over time, suggesting

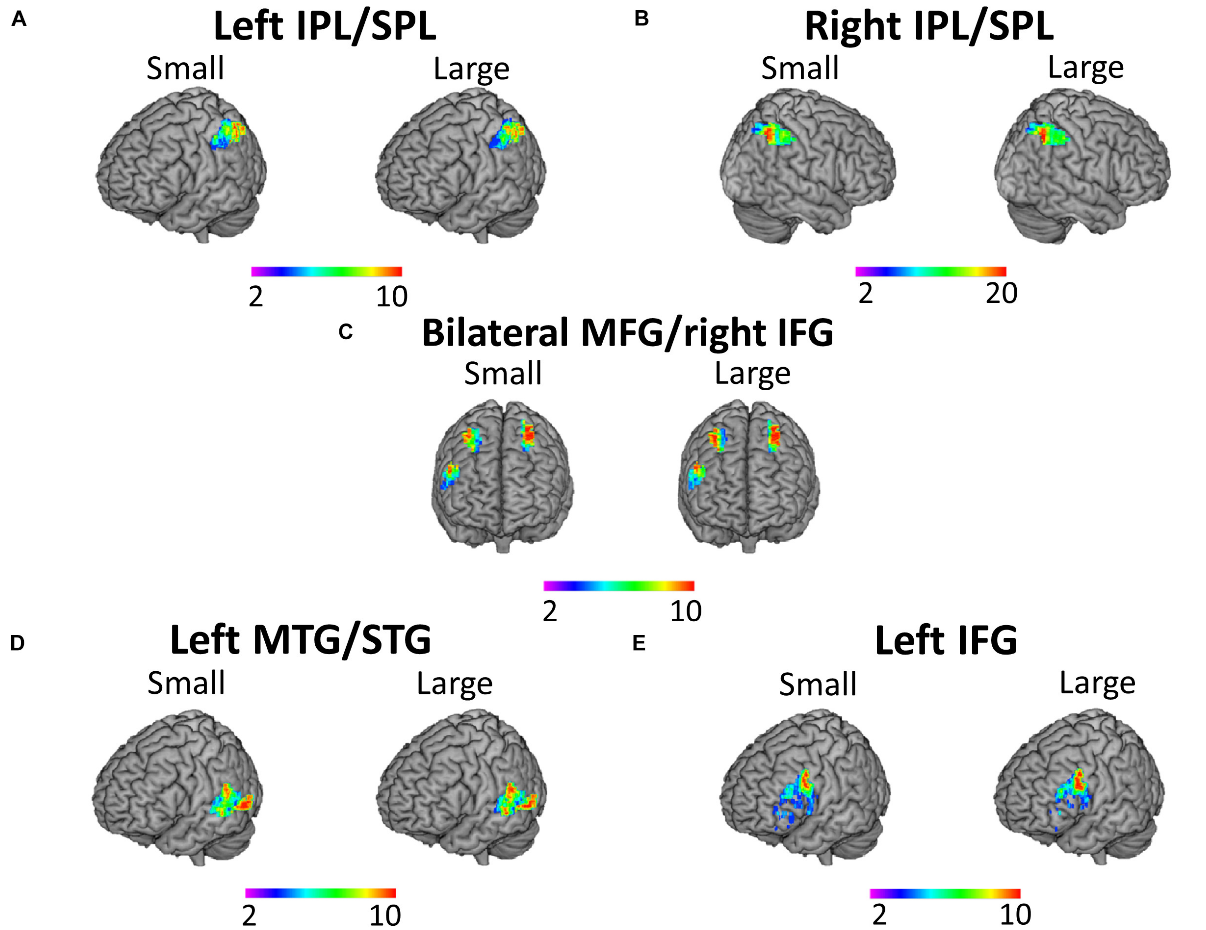


FIGURE 3 | Cluster overlap across participants for activation during subtraction problem-solving. Cluster overlap across participants for the 100 voxels showing maximal activation separately for small and large subtractions in quantity representation ROIs in **(A)** left and **(B)** right IPL/SPL; **(C)** access to quantity representation ROIs in bilateral MFG/right IFG; **(D)** verbal representation ROI in left MTG/STG; and **(E)** access to verbal representation ROI in left IFG. Color bar shows the number of participants showing overlap, from 2 participants shown in purple/blue colors to 10 participants shown in yellow/red colors for all the regions except right IPL/SPL, in which case yellow/red colors indicate 20 participants.

that children continue to rely on procedures. Critically, we expected to see decreases in bilateral MFG/right IFG over time, suggesting that procedures become more automatic, a central claim of the Schema-based hypothesis. In the second case, illustrated in arrow B in **Figure 5**, we expected to see decreases in bilateral IPL/SPL and increases in left MTG/STG over time. It is possible that this pattern is accompanied by increases in left IFG activation over time, given that the implementation of retrieval strategies might be effortful in the early stages (Geary et al., 1996a; i.e., Fact-retrieval hypothesis). Finally, there is a third possibility, illustrated in arrow C in **Figure 5**. Evidence suggests that by 10 years of age retrieval is the dominant strategy to solve single-digit arithmetic problems (Ashcraft and Fierman, 1982), so it is possible that children may have already shifted toward retrieval by the time they were scanned at time 1. In this case, we expected left MTG/STG activation at time 1 to predict subtraction fluency gains and children to show increases in temporal cortex activation over time. These findings would suggest that children continue to build their storage of subtraction facts in long-term memory. This pattern might be accompanied by decreases in left IFG over time, indicating that retrieval becomes less effortful as the representations in long-term memory become more robust (Prado et al., 2014).

Whole Brain Analysis

In order to investigate the effects outside our ROIs, we ran a two-sample *t-test* comparing brain activity between improvers and non-improvers at the whole brain (i.e., after excluding ROIs). Following the ROI analysis logic, we focused on (a) brain activation at time 1 by looking at the contrast "large subtractions vs. control at time 1" and (b) changes in brain activation over time by looking at the contrast "large subtractions vs. control time 2 – time 1". Statistical significance for the whole brain was defined using 3dClustSim. A cluster size of 175 voxels was needed for whole brain significance (ACF values = 0.45, 4.57, 11.14) at a cluster-wise threshold of $p = 0.05$ and a voxel-wise threshold of $p = 0.005$.

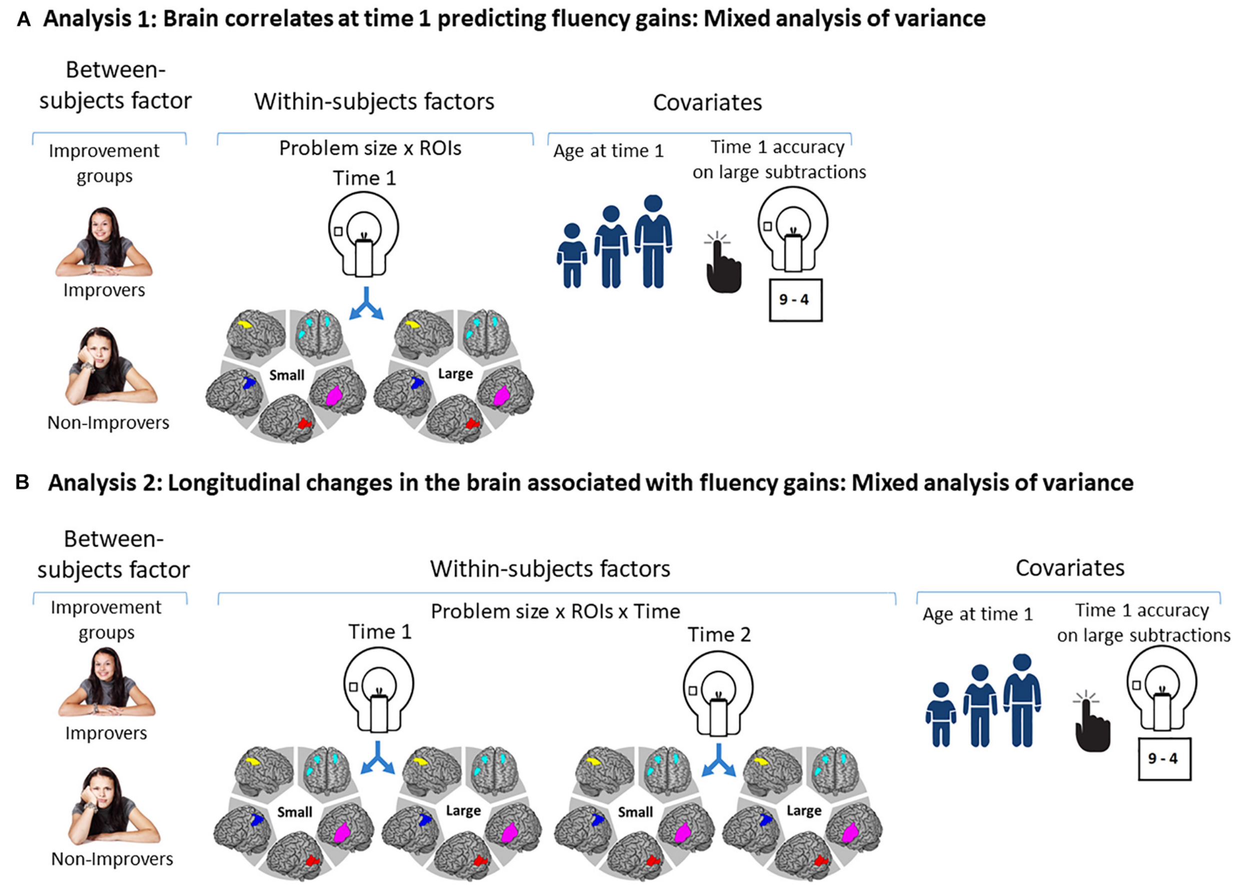


FIGURE 4 | Illustration of the factors included in the statistical analyses. **(A)** Illustration of the between-subjects factors, within-subjects factors, and covariates included in the mixed ANOVAs calculated to study whether improvement groups differed in the brain regions they engaged to solve subtraction problems at time 1. **(B)** Illustration of the between-subjects factors, within-subjects factors, and covariates included in the mixed ANOVAs performed to study the changes in brain activation associated with longitudinal gains in subtraction fluency.

RESULTS

Localizer Tasks Behavioral Results

We calculated accuracy and RTs (for correctly solved trials) for the rhyming and numerosity judgment localizer tasks, for the participants whose data were used to define ROIs (i.e., $n = 40$; see section "Regions of Interest Definition" for more information). Repeated-measures ANOVAs were performed separately for accuracy and response times, and separately for each localizer task. For the numerosity task, we entered *Difficulty* as the within-subjects factor, which referred to the distance between the number of dots to be compared: easy (12 vs. 36), medium (18 vs.36), and hard (24 vs. 36). As for the rhyming judgment task, we included *Conflict*, which referred to whether orthography was consistent (i.e., non-conflicting) or inconsistent (i.e., conflicting) with the correct answer and *Rhyming*, referring to whether the pair of words rhymed or not, as the within-subject factors. *Post hoc* tests, using Bonferroni correction, were calculated when an effect was found significant.

As for the numerosity judgment task, we found a main effect of *Difficulty* for accuracy [$F(2,78) = 6.07$, $p = 0.004$, *partial* $\eta^2 = 0.14$], showing that accuracy was highest for the easy condition (mean = 90.92, SEM = 1.41), lowest for the hard condition (mean = 86.05, SEM = 1.77), and intermediate for the medium condition (mean = 88.55, SEM = 1.78). The *Difficulty* effect was also significant for the response time analysis [$F(2,78) = 14.37$, $p < 0.001$, *partial* $\eta^2 = 0.27$], and showed fastest response times for the easy condition (mean = 976 ms, SEM = 237), slowest response times for the hard condition (mean = 1061 ms, SEM = 258), and intermediate response times for the medium condition (mean = 1018 ms, SEM = 245).

Regarding the rhyming judgment task, the accuracy analysis showed a main effect of *Rhyming* [$F(1,39) = 28.58$, $p < 0.001$, *partial* $\eta^2 = 0.42$]. Children were more accurate for pairs that rhymed (mean = 90.77, SEM = 1.38) as compared to pairs that did not rhyme (mean = 70.83, SEM = 3.73). The same effect was shown in response times [$F(1,36)$[8] $= 39.95$, $p < 0.001$, *partial* $\eta^2 = 0.53$], with children being faster on

[8]The difference in degrees of freedom is due to 3 participants having no correct responses for one of the four conditions, so response times could not be calculated, resulting in missing data.

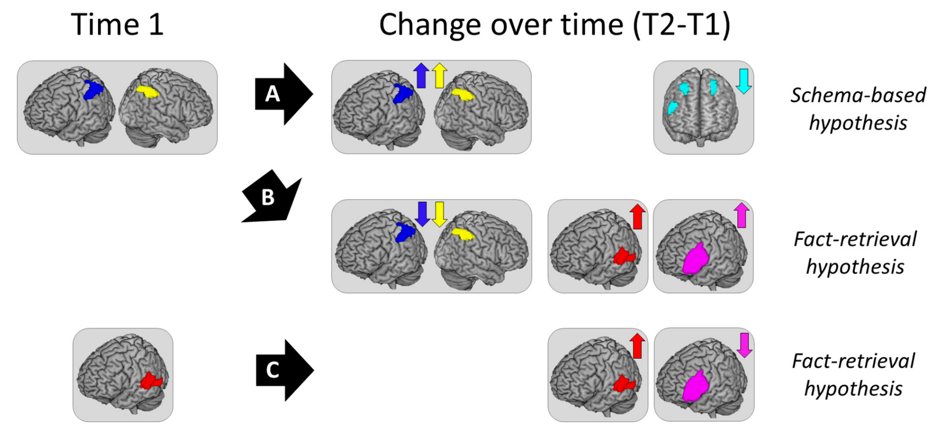


FIGURE 5 | Illustration of the predictions for each hypothesis. **(A)** Arrow indicating the predictions for the Schema-based hypothesis: Initial activation of bilateral IPL/SPL at time 1 followed by increases in brain activation in this region over time, suggesting that children continue to rely on procedures to become more fluent in subtractions. In addition, decreases in bilateral MFG/right IFG would indicate that the implementation of these procedures become more automatic over time, a central claim of the Schema-based hypothesis. **(B)** Arrow indicating a main prediction that would support the Fact-retrieval hypothesis: Initial activation of bilateral IPL/SPL at time 1 is followed by decreases in brain activation in this region and by increases in left MTG/STG activation over time, suggesting that children shift from procedures to retrieval of the solution from long-term memory, a core claim of the Fact-retrieval hypothesis. These changes in brain activation might be accompanied by increases in left IFG over time, considering that the implementation of the retrieval strategy might be effortful in its early stages (Geary et al., 1996a). **(C)** Arrow indicating an alternative prediction that would support the Fact-retrieval hypothesis. Given children 10 years of age use the dominant strategy of retrieval to solve single-digit arithmetic problems (Ashcraft and Fierman, 1982), it may be the case that children have already shifted toward retrieval by the time they were scanned at time 1, which would be consistent with the finding that brain activation in left MTG/STG at time 1 predicts subtraction fluency gains over time. This alternative hypothesis also predicts that brain activation in temporal cortex would increase over time, suggesting that children continue to build their long-term storage of subtraction facts. It is also possible that this change is accompanied by decreases in left IFG activation over time, indicating that retrieval becomes less effortful as the representations in long-term memory become more robust (Prado et al., 2014).

rhyming pairs (mean = 1185 ms, SEM = 42) than non-rhyming ones (mean = 1364 ms, SEM = 48). We also found a main effect of *Conflict* for accuracy [$F(1,39) = 64.65$, $p < 0.001$, *partial* $\eta^2 = 0.62$], with children being more accurate for non-conflicting (mean = 89.23, SEM = 11.67) than for conflicting pairs (mean = 72.37, SEM = 17.50). The same main effect of *Conflict* was found for response times [$F(1,36) = 18.86$, $p < 0.001$, *partial* $\eta^2 = 0.34$], with children taking longer to respond to conflicting (mean = 1315 ms, SEM = 293) than to non-conflicting pairs (mean = 1223 ms, SEM = 246). The *Rhyming* × *Conflict* interaction was also significant for accuracy [$F(1,39) = 24.17$, $p < 0.001$, *partial* $\eta^2 = 0.38$]. While the comparisons across all conditions were significant (all *p*-Values below.005), the interaction showed that the non-rhyming condition with conflicting orthography (O + P−) was the hardest. The *Rhyming* × *Conflict* interaction was also significant for response times [$F(1,36) = 8.23$, $p = 0.007$, *partial* $\eta^2 = 0.19$]. The interaction was due to a significant difference between the conflicting and non-conflicting conditions among the non-rhyming pairs ($p < 0.001$; O + P− and O-P-), but a non-significant difference between conflicting and non-conflicting conditions among the rhyming pairs ($p = 0.22$; O + P+, and O−P +).

Subtraction Task Behavioral Results

Whole Sample Performance

We calculated accuracy and means of RTs (for correctly solved trials) separately for small and large subtractions, for every participant.

We calculated a repeated measures ANOVA for accuracy including Time (i.e., time 1, time 2) and Problem size (i.e., small, large) as within-subjects factors. *Post hoc* tests, using Bonferroni correction, were calculated when an effect was found significant. We found a main effect of Time [$F(1,45) = 23.15$, $p < 0.001$, *partial*$\eta^2 = 0.34$] and a main effect of Problem size [$F(1,45) = 25.67$, $p < 0.001$, *partial*$\eta^2 = 0.36$], but no Time × Problem size interaction [$F(1,45) = 0.49$, $p = 0.49$, *partial*$\eta^2 = 0.01$]. The Time effect showed that, across problem sizes, children were more accurate at time 2 (mean = 89.53, SEM = 1.24) as compared to time 1 (mean = 80.50, SEM = 2.08; $p < 0.001$). The Problem size effect showed that, across time points, children were more accurate solving small (mean = 87.70, SEM = 1.28; $p < 0.001$) as compared to large (mean = 82.33, SEM = 1.74) subtractions.

We then calculated a repeated measures ANOVA for means of RTs including Time (i.e., time 1, time 2) and Problem size (i.e., small, large) as within-subjects factors. We found a main effect of Time [$F(1,45) = 49.33$, $p < 0.001$, *partial*$\eta^2 = 0.52$], and a main effect of Problem size [$F(1,45) = 44.46$, $p < 0.001$, *partial*$\eta^2 = 0.50$], but no Time × Problem size interaction [$F(1,45) = 0.005$, $p = 0.94$, *partial*$\eta^2 = 0.00$]. The Time effect showed that, across problem sizes, children were faster at time 2 (mean = 1049 ms, SEM = 52) as compared to time 1 (mean = 1313 ms, SEM = 54; $p < 0.001$). The Problem size effect showed that, across time points, children were faster to solve small (mean = 1113 ms, SEM = 48; $p < 0.001$) as compared to large (mean = 1250 ms, SEM = 52) subtractions.

Improvement Groups' Performance

We then explored children's performance in large subtractions depending on improvement groups (see section "Improvement Groups" for a description of how groups were formed). This confirmatory analysis was carried out to test whether groups showed the expected pattern of behavioral changes over time.

We calculated a repeated-measures ANOVA for accuracy entering Time (Time 1; Time 2) as the within-subjects factor and Improvement groups (improvers, non-improvers) as the between-subjects factor. The same ANOVA was calculated for RTs.

As for accuracy, we found a significant main effect of Time [$F(1,44) = 16.72$, $p < 0.001$, $partial\eta^2 = 0.28$], but no Time × Improvement group interaction [$F(1,44) = 0.62$, $p = 0.44$, $partial\eta^2 = 0.01$]. The main effect of Time showed that, regardless of improvement group, all children became more accurate [$t(45) = 5.66$, $p < 0.001$]. More detailed information about groups' performance is given in **Table 3**. The main effect of Time did not reach significance when age at time 1 was entered as a covariate in the ANOVA [$F(1,43) = 1.71$, $p = 0.20$, $partial\eta^2 = 0.04$].

Regarding RTs, the main effect of Time [$F(1,44) = 73.97$, $p < 0.001$, $partial\eta^2 = 0.63$] was significant. As expected, based on the definition of the improvement groups, the Time × Improvement group interaction was also significant [$F(1,44) = 59.74$, $p < 0.001$, $partial\eta^2 = 0.58$]. The interaction showed that the improvers had a significant decrease in RTs over time [$t(22) = 11.17$, $p < 0.001$], whereas the non-improvers did not [$t(22) = 0.64$, $p = 0.53$]. Groups differed in RTs at time 2 [$t(44) = -4.60$, $p < 0.001$], but not at time 1 [$t(44) = 0.33$, $p = 0.75$]. More detailed information about groups' performance is given in **Table 3**. **Figure 6** shows the changes over time in RTs for improvers and non-improvers. Results were consistent if age at time 1 was entered as a covariate in the ANOVA (i.e., Time × Improvement group interaction: $F(1,43) = 61.06$, $p < 0.001$, $partial\eta^2 = 0.59$).

fMRI Results

Improvers Showed a Larger Neural Problem Size Effect in Bilateral Parietal Cortex at Time 1

The analysis of brain activation at time 1 showed a significant ROI × Problem size × Improvement groups interaction [$F(3,138) = 2.66$, $p = 0.04$, $partial\eta^2 = 0.06$, Greenhouse-Geisser $\varepsilon = 0.82$][9]. We explored the three-way interaction with pairwise comparisons using Bonferroni correction to control for multiple comparisons. This analysis showed differences in the left IPL/SPL [$t(22) = -4.41$, $p = 0.001$] and right IPL/SPL [$t(22) = -3.57$, $p = 0.01$] between small and large subtractions only for improvers. **Figure 7** shows the differences in brain activation in the left (A) and right (B) IPL/SPL between small and large subtractions for the improvers and non-improvers groups. The two group did not differ in bilateral MFG/R IFG ($p = 0.47$), left MTG ($p = 0.26$), or left IFG ($p = 0.17$) at time 1. The non-improvers showed no neural problem size effect in any of the ROIs (all p-Values above 0.25).

[9]Results remained significant if the top 50 [$F(3,136) = 3.00$, $p = 0.03$, $partial\eta^2 = 0.07$, Greenhouse-Geisser $\varepsilon = 0.81$] or top 200 [$F(3,137) = 3.60$, $p = 0.01$, $partial\eta^2 = 0.08$, Greenhouse-Geisser $\varepsilon = 0.82$] voxels were selected instead of the top 100.

TABLE 3 | Performance on large subtractions solved inside the scanner.

	Whole ($n = 46$)	Improvers large ($n = 23$)	Non-improvers large ($n = 23$)
Accuracy T1	77.5 (17.1)	79.1 (15.7)	75.9 (18.6)
Accuracy T2	87.2 (10.7)	86.9 (11.5)	87.4 (10.0)
Accuracy change	9.7 (16.0)	7.8 (13.7)	11.5 (18.1)
RTs T1	1382 (387)	1401 (441)	1363 (333)
RTs T2	1117 (389)	898 (350)	1336 (294)
RTs change	−265 (317)	−503 (216)	−27 (202)

Means of response times (RTs; in milliseconds), and accuracy (standard deviation in parentheses) at time 1 (T1) and time 2 (T2) and change (time 2 − time 1) for the whole sample (n = 46), for improvers (n = 23), and for non-improvers (n = 23).

Improvers Decreased Activation for Large Subtractions in Both Parietal and Frontal ROIs Over Time

The ANOVA showed a Time × ROI × Problem size × Improvement groups interaction [$F(3,126) = 2.76$, $p = 0.04$, $partial\eta^2 = 0.06$, Greenhouse-Geisser $\varepsilon = 0.75$].[10] Pairwise comparisons using Bonferroni correction showed

[10]Results remained significant if the top 50 [$F(3,125) = 3.10$, $p = 0.03$, $partial\eta^2 = 0.07$, Greenhouse-Geisser $\varepsilon = 0.75$] or the top 200 [$F(3,125) = 2.75$, $p = 0.046$, $partial\eta^2 = 0.06$, Greenhouse-Geisser $\varepsilon = 0.75$] voxels were selected instead of the top 100.

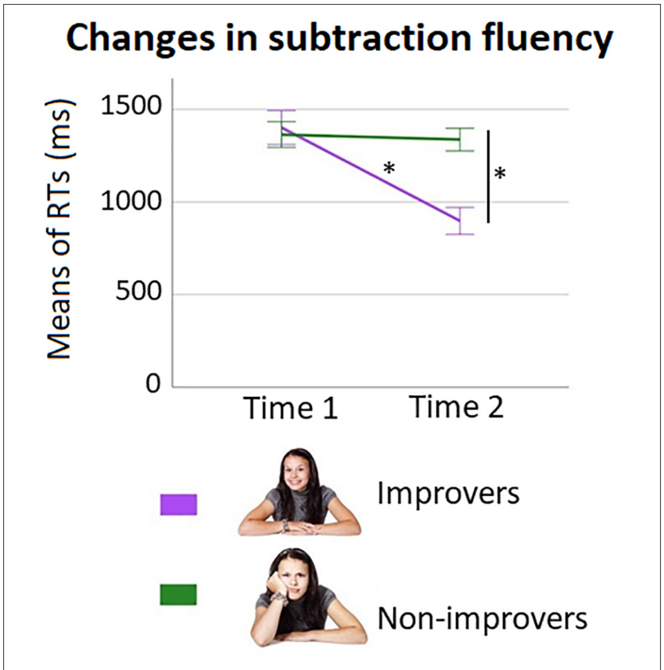


FIGURE 6 | Improvement groups changes in response times (RTs) over time. Changes in RTs (in milliseconds) over time to solve large subtractions separately for improvers and non-improvers. Error bars show standard error of the mean. Asterisks indicate significant differences between groups at time 1 and between time points for improvers.

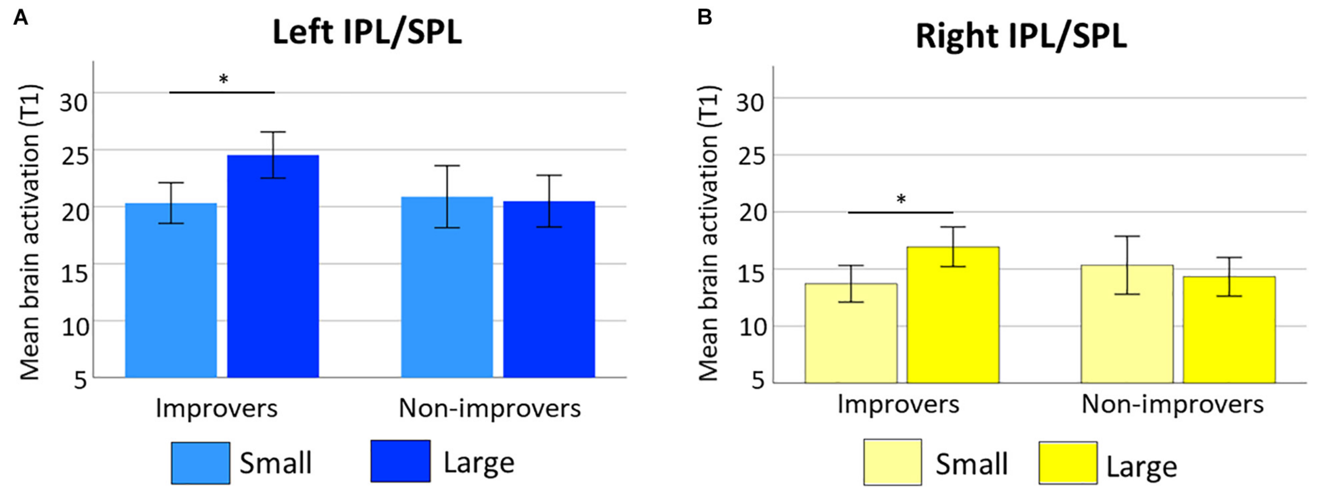


FIGURE 7 | Neural problem size effect in bilateral parietal cortex at time 1 for improvement groups. Bar chart show brain activation in **(A)** left IPL/SPL and **(B)** right IPL/SPL for small (i.e., lighter colors) and large (i.e., darker colors) subtractions, for improvers and non-improvers. Error bars show standard error of the mean. Asterisks indicate significant differences between problem sizes in left and right IPS for improvers.

a different pattern of changes in brain activation over time depending on problem size. For small subtractions, both groups showed significant decreases in brain activation over time in all ROIs (all p-Values below.02). As for large subtraction problems, improvers showed a significant decrease over time in all ROIs (all p-Values equal or below.001), whereas non-improvers showed significant decreases over time only for left MTG/STG ($p = 0.005$), but not for left IPL/SPL ($p = 0.40$), right IPL/SPL ($p = 0.13$), bilateral MFG/right IFG ($p = 0.07$) or left IFG ($p = 0.054$). **Figure 8** illustrates changes over time in brain activation for large subtraction problems for improvers (i.e., plain bars) and non-improvers (i.e., patterned bars).

Evidence for the Efficiency of Numerical Procedures: A Exploratory Analysis of the Improvers

We aimed to further explore the idea of efficiency of numerical procedures by more closely looking at the improvers group. Considering our finding, showing decreased bilateral MFG/right IFG activation over time for improvers, it would be reasonable to expect greater decreases in these regions for children becoming faster over time, even among the improvers ($n = 23$), providing further evidence for the automaticity in the implementation of procedures. To this aim, participants in the improvers group were split into two subgroups: *slower* improvers ($n = 11$) and *faster* improvers ($n = 12$), based on the same procedure used to define the improvers vs. non-improvers and described in section "Improvement Groups". As shown in **Figure 9A**, the *slower* improvers group [$t(10) = 7.50$, $p < 0.001$] and the *faster* improvers group [$t(11) = 10.77$, $p < 0.001$] significantly decreased in response times over time, but they differed in how fast they solve problems at time 2 [$t(21) = 2.66$, $p = 0.01$].

We also ran a student *t-test* comparing bilateral MFG/right IFG brain activation between *slower* and *faster* improvers at each time point. As shown in **Figure 9B**, results showed that groups differed in brain activation at time 2 [$t(21) = -2.27$, $p = 0.03$], but not at time 1 [$t(21) = -1.16$, $p = 0.26$], with the *faster* subgroup showing less bilateral MFG/right IFG activation than their *slower* counterparts at time 2. Both the *slower* [$t(10) = 3.98$, $p = 0.003$] and the *faster* [$t(11) = 5.32$, $p < 0.001$] improvers subgroups significantly decreased brain activation in this area over time.

Whole Brain Results

Three clusters (shown in **Supplementary Figure 1**) reached significance for the contrast "large subtractions vs. control time 2 – time 1", showing greater activation for non-improvers as compared to improvers. More specific information about these clusters is provided in **Table 4**. No cluster reached significance for the contrast "large subtractions vs. control time 1".

DISCUSSION

Despite the crucial role that mathematics plays in our society for personal and professional development, and the importance that developing math fluency has in the acquisition of more advanced mathematics (Geary, 1994; Price et al., 2013), the neurocognitive mechanisms associated with improvement in behavioral fluency are poorly understood. While there is consensus in the literature that children show a shift toward retrieval for operations such as multiplication (Ashcraft, 1982; Dehaene and Cohen, 1995), it is not yet clear how fluency is achieved in subtraction. Two hypotheses have been formulated to explain subtraction fluency development. According to the Fact-retrieval hypothesis, children become fluent in single-digit subtractions by shifting from procedures to the retrieval of the solutions from declarative long-term memory (Ashcraft, 1982; Siegler, 1987). The Schema-based hypothesis, on the other hand, claims that children achieve subtraction fluency by means of procedures that

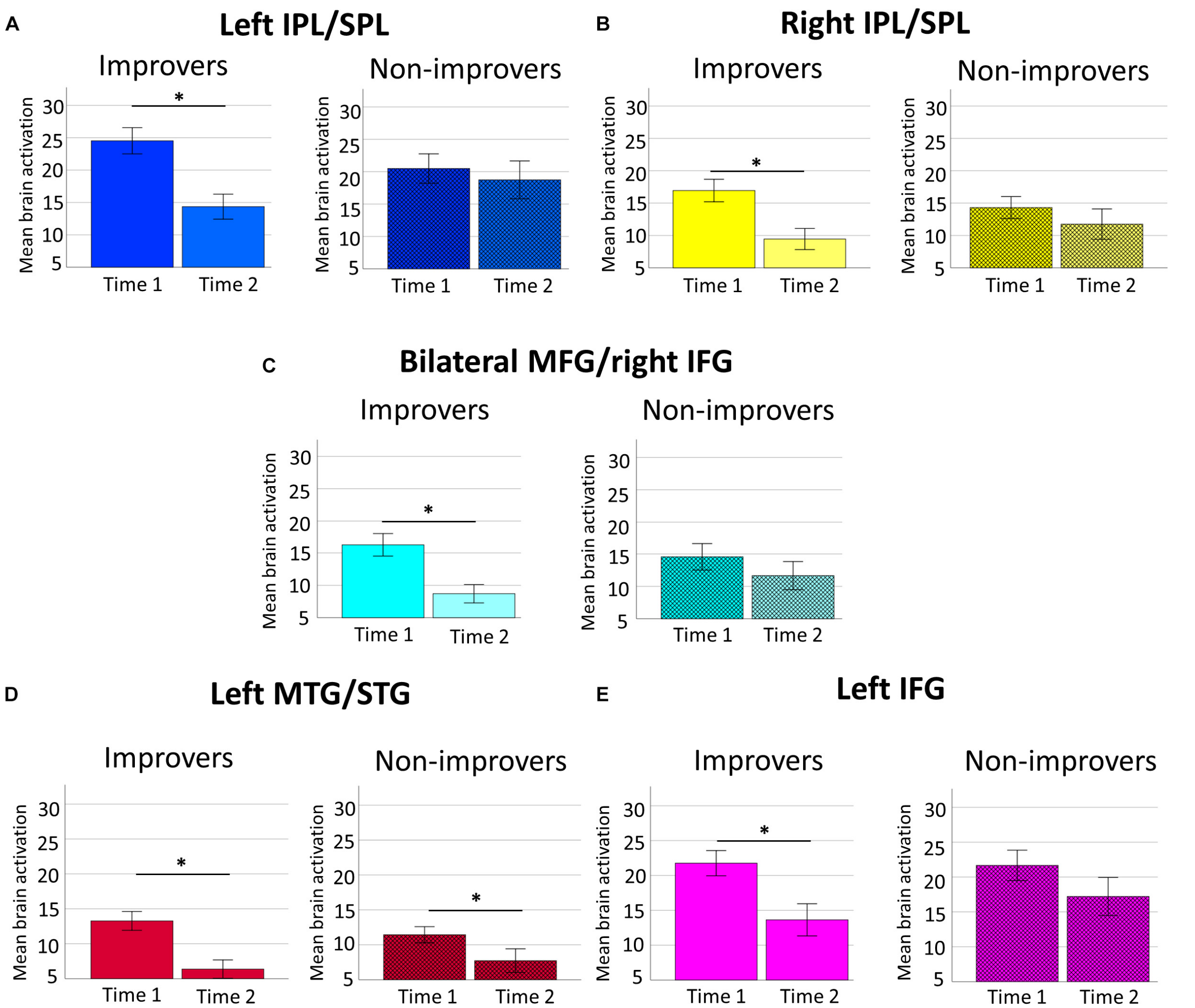


FIGURE 8 | Changes in brain activation over time for large subtractions for improvers and non-improvers. Changes in brain activation in **(A)** left IPL/SPL, **(B)** right IPL/SPL, **(C)** bilateral MFG/right IFG, **(D)** left MTG/STG, and **(E)** left IFG for children showing improvement (i.e., plain bars) and non-improvement (i.e., patterned bars). Error bars show standard error of the mean. Asterisks indicate significant differences between time points in all ROIs for improvers and in left MTG/STG for non-improvers.

become automatic over development (Baroody, 1983; Fayol and Thevenot, 2012). Given that both hypotheses make the same predictions regarding changes in RTs but differ in the mechanisms considered to be responsible for that change, and given that automatic processes seem to be easily confounded with and reported as retrieval (Fayol and Thevenot, 2012), neither RTs nor self-reported measures have been able to adjudicate between these two hypotheses. Within this context, fMRI can help by investigating (a) whether engagement of verbal or quantity brain areas early on (i.e., time 1) predict longitudinal gains in subtraction fluency; (b) whether longitudinal fluency gains are associated with changes in verbal or quantity brain activation. Importantly, the aim of this study was to assess differences in the neurocognitive mechanisms recruited by children differing in fluency but showing similar levels of accuracy on the experimental task.

Modulation of Parietal Cortex by Problem Size at Time 1 Predicts Longitudinal Subtraction Fluency Improvement

When examining the role of brain activation at time 1 in predicting longitudinal gains in subtraction fluency we found that improvers showed a larger neural problem size effect in bilateral IPS at time 1, with greater activation for large subtractions as compared to small ones. These results are consistent with Prado et al. (2014)'s cross-sectional evidence showing grade-related increases in parietal cortex for solving subtractions with more years of math instruction. Prado

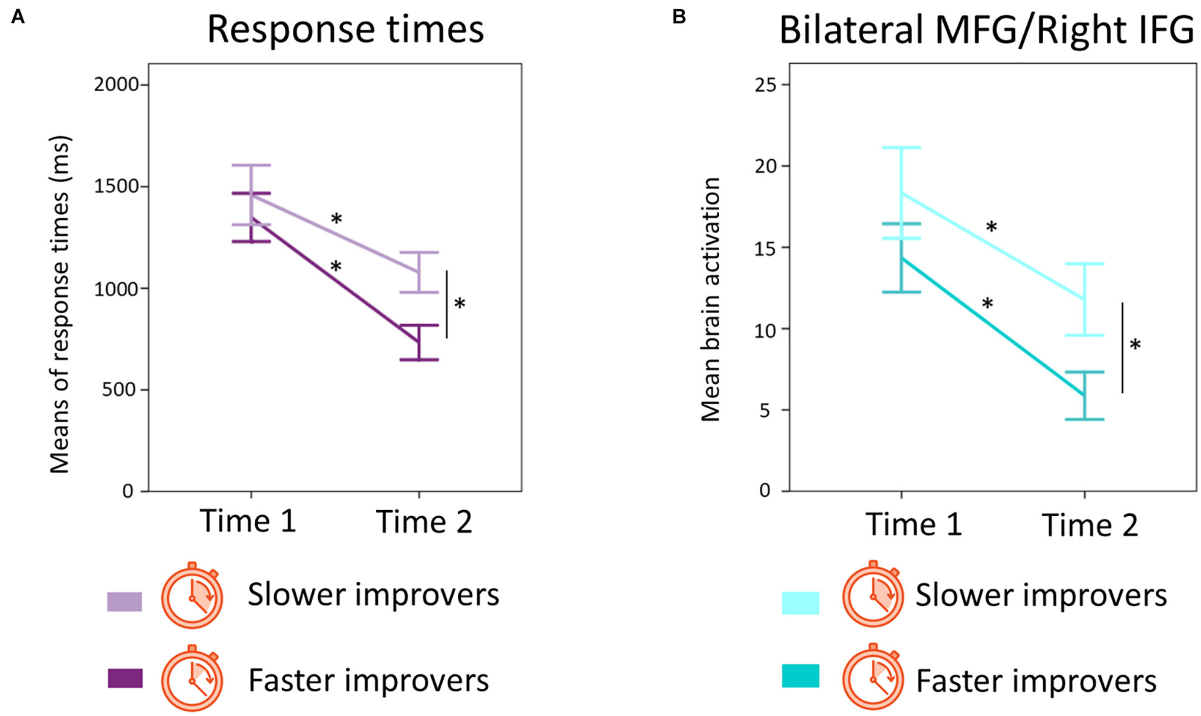


FIGURE 9 | Changes in RTs and changes in bilateral MFG/right IFG activation for the subgroup of improvers that ended up being faster (i.e., *faster improvers*) or slower (i.e., *slower improvers*) at time 2. **(A)** Changes over time in response times for the subgroup of faster and slower improvers. Error bars show standard error of the mean. Asterisks confirm significant differences between time points for both subgroups and significant differences in response times between subgroups at time 2. **(B)** Changes over time in bilateral MFG/right IFG activation for the subgroups of faster and slower improvers. Error bars show standard error of the mean. Asterisks indicate significant differences between time points for both subgroups and significant differences in brain activation between subgroups at time 2.

et al. (2014)'s and our results both support the involvement of quantity but not verbal regions for subtraction learning. However, there were some differences between studies. First, the covariate of interest in our study was how much children improved from time 1 (i.e., sample in Prado's papers) to time 2, with age being controlled for. In contrast, Prado et al. (2014) study included grade (i.e., second through eighth), which is highly correlated with age, as the predictor of interest. Second, while Prado et al. (2014) found the effects only for small subtractions, we found them for large subtractions. Third, while Prado et al. (2014) found the effects in the right posterior superior parietal lobule (PSPL), we found them in bilateral IPS. While the IPS plays a role in representing quantities (Dehaene et al., 2003; Piazza et al., 2004), the PSPL has been associated with visuo-spatial attentional processes in children (Krinzinger, 2011) and adults (Simon et al., 2002). Previous results have suggested overlapping patterns of activity in PSPL for addition and subtractions and shifts of visuo-spatial attention (Knops et al., 2009), like the ones needed to estimate the position along a mental number line (Berteletti et al., 2014). Prado et al. (2014) concluded that these visuo-spatial shifts seemed to be sufficient for solving small subtractions, while solving large ones would require more involvement of quantity mechanisms in IPS. Our results confirm Prado's predictions by showing that engaging these quantity mechanisms in IPS, early on, explained longitudinal gains in subtraction fluency.

Interpretation of the Decreases in Parietal Activation Over Time as Supporting the Schema-Based Hypothesis

The fact that parietal cortex at time 1 was the only ROI that predicted subtraction fluency improvement does not distinguish between the Schema-based and Retrieval-based hypotheses. It could be the case that this early parietal engagement, suggesting procedural use, is replaced by the retrieval of the solution from memory, in which case we should see a shift in brain activation from parietal to temporal regions. We hypothesized that this could be accompanied by increases in left IFG, given that the implementation of retrieval strategy might be effortful in young children (Geary et al., 1996a). Alternatively, it might be the case that the use of procedures is not replaced, but becomes more efficient over time, in which case we would see increases in parietal cortex over time, suggesting that children continue to rely on procedures to develop their fluency. This should be accompanied by decreases in bilateral MFG/right IFG over time, suggesting that procedures become more automatic, a core claim of the Schema-based hypothesis.

TABLE 4 | Whole brain results.

K	MNI coordinate			Z-score	∼BA	Anatomical region
	X	Y	Z			
176	−2	−81	22	3.93	17	left cuneus and left calcarine
	−2	−91	18	3.48		
	2	−83	6	2.78		
194	32	−33	38	3.70	2, 3, 40	right postcentral and right supramarginal gyrus
	40	−35	46	3.49		
	44	−23	42	3.08		
215	−38	−53	6	3.93	37	left middle occipital cortex
	−36	−65	2	3.63		
	−36	−67	22	3.50		

Clusters showing significant activation at the whole brain level for large subtraction problems as compared to the control condition at time 2 as compared to time 1 for the non-improvers as compared to the improvers.

The analysis of longitudinal changes in brain activation showed that children who improved in subtraction fluency decreased activation in bilateral IPL/SPL over time. Previous evidence has suggested that *less* activation for a given level of proficiency represents more efficient use of certain brain regions (Prat et al., 2007; Prat and Just, 2011). Several fMRI studies have found that more skilled or highly trained individuals show less brain activation as compared to controls (Rypma and D'Esposito, 1999; Krings et al., 2000; Welcome and Joanisse, 2012). In addition, decreased activation in different brain regions has been found after practice with visuomotor association tasks (Büchel et al., 1999), visuospatial WM (Garavan et al., 2000), verbal WM (Hempel et al., 2004), Tower of London (Beauchamp et al., 2003), or counting Stroop tasks (Bush et al., 1998) in which participants became faster with training. In the field of math cognition, previous work has shown decreased brain activation for perfect performers (i.e., 100% accuracy) as compared to imperfect performance (i.e., 78%-96% accuracy; Menon et al., 2000), suggesting that after a certain level of expertise is achieved, the brain can achieve the same results with fewer resources. Within this context, and considering that the reductions in parietal cortex activation were unique to the improvers group, we interpret our findings as showing that gains in subtraction fluency is associated with a more efficient recruitment of parietal cortex by calculation procedures. The fact that non-improvers continue to engage parietal cortex over time is consistent with previous evidence showing greater bilateral parietal activation for children with developmental dyscalculia solving a subtraction task as compared to an addition task (Rosenberg-Lee et al., 2015). This finding is also consistent with evidence showing that 8 weeks of one-to-one math tutoring resulted in significant reductions in overactivation of bilateral IPS (among other regions) in children with math learning disabilities (Iuculano et al., 2015).

Decreases in Bilateral MFG/Right IFG Supports the Schema-Based Hypothesis

Our finding of a reduction in brain activation in bilateral MFG/right IFG supports the Schema-based hypothesis and suggests that processes occurring in these areas become more automatic over time. Decreases in bilateral parietal and frontal regions were interpreted as evidence for procedures becoming more efficient in a training study with adults that found untrained subtractions engaged bilateral IPS and bilateral IFG as compared to trained ones (Ischebeck et al., 2006). Less activation in MFG over development has also been observed in a cross-sectional study in 8—19-year-old children solving additions and subtractions (Rivera et al., 2005), in a longitudinal study of 6th to 7th-grade children solving two-digit subtractions (Artemenko et al., 2018), and in adults solving arithmetic problems as compared to children (Kawashima et al., 2004; Kucian et al., 2008). Our finding adds to this evidence by showing that frontal regions involved in quantity processing, as identified with a numerosity judgment localizer task, decreased activation with improvement in subtraction fluency and support the hypothesis of increased automaticity in accessing procedures.

Underlying Mechanisms Explaining the Automaticity of Procedures

Our findings support previous studies suggesting that, at least for subtractions, developing fluency involves procedures becoming more automatic over time (Baroody, 1983; LeFevre et al., 2006; Fayol and Thevenot, 2012). Using a priming paradigm, Fayol and Thevenot (2012) tested whether solving additions, subtractions, and multiplications mobilized a procedural component or were solved by retrieval. They tested whether procedures were pre-activated as soon as individuals see a sign (i.e., +, -, x), presented before the arithmetic problem indicating the upcoming operation. They found that solving additions and subtractions was facilitated when the operation sign was presented 150 ms before the operands and that this effect was operation-specific. They inferred that abstract procedures were primed by the presentation of the sign, subsequently helping with solving the problems. The presentation of the multiplication sign had no facilitation effect on solving the problems, confirming the hypothesis that they did not rely on procedures. Moreover, subtractions were not solved slower than multiplications, suggesting that procedures could be as fast as retrieval. In a similar study, Mathieu's et al. (2016) presented the first operand and the operator in the center of the screen, while the second operand was presented either in the left or the right of the screen. They found that additions were solved faster when the second operand appeared to the right of the screen whereas subtractions were solved faster when the operand was presented to the left. No effect was found for multiplication. They interpreted these findings as suggesting that solving additions and subtractions activated procedures consisting of rightward and leftward shifts of attention, respectively, along a mental number line. Furthermore, in a study of the neural correlates of these effects, Mathieu's et al. (2018) found greater activation in brain regions supporting the orientation of spatial attention, including right posterior superior parietal lobule (PSPL), when participants were presented with the "+" sign as compared to the "x" one. They interpreted that the operation-priming effect shown by Fayol and Thevenot (2012) was due to arithmetic symbols evoking spatial mechanisms that would, in turn, lead to facilitation of performance for that operation.

While previous studies from our lab have interpreted grade-related findings in the PSPL for small subtractions as suggesting visuo-spatial attentional shifts, this interpretation seems less likely to explain our findings in the IPS, a brain region well known for its role in quantity representation (Dehaene et al., 2003; Piazza et al., 2004). We believe that reliance on quantity-based procedures becomes more automatic because these representations in parietal cortex are refined over development (Suárez-Pellicioni and Booth, 2018). Several studies have shown that with experience to symbolic mathematics, children develop more precise representation of quantities (Ansari, 2008; Mussolin et al., 2014; Matejko and Ansari, 2016). If quantity representations are more refined, children are better able to implement calculation procedures more efficiently, requiring less parietal activation. More precise quantity representations would also explain the decreases in bilateral MFG/right IFG regions over time, suggesting calculation procedures become less effortful.

The Case of Addition: An Ongoing Debate

A consensus has not yet been reached regarding whether the Fact-retrieval or the Schema-based hypotheses better explain arithmetic fluency development for arithmetic problems involving addition. As mentioned above, Fayol and Thevenot (2012), Mathieu's et al. (2018, 2016) results suggest that solving both addition and subtraction problems rely on procedures. Barrouillet and Thevenot (2013), showed that response times monotonically and linearly increased when addends were incremented by one, a finding they argued is not consistent with retrieval use, but rather points to adults relying on fast procedures to solve additions. Uittenhove et al. (2016) also argued that it was difficult to interpret the high variability in response times to addition problems resulting from a one-step direct retrieval process. Finally, Thevenot et al. (2016) aimed to challenge previous evidence suggesting that by 10 years old children already rely on retrieval to solve single-digit arithmetic problems (Ashcraft and Fierman, 1982). Their analysis of 10-year-old children's response time patterns to a single-digit addition production task was compatible with shifting from slow to fast counting procedures but not with a shift toward retrieval.

A very recent study used EEG to try to clarify between the fact-retrieval and the schema-based hypotheses by administering adults a single-digit addition and multiplication production task. Their analyses of theta, lower alpha, and upper alpha frequencies showed higher evidential strength for similar EEG activity between very small additions (i.e., operands between 1 and 4) and multiplication problems, suggesting that very small additions are solved through fact retrieval, and supporting the Fact-retrieval hypothesis (Grabner et al., 2020). Other studies investigating subtraction problem solving using fMRI have reported evidence suggesting that additions are solved through retrieval, with the hippocampus playing a potentially important role in memory formation for these facts (Cho et al., 2012). Using a multivariate analysis, Cho et al. (2011) found differences in neuronal activity patterns between 7- to 9-year-old children that were classified as retrievers vs. counters when solving a single-digit addition task, with the highest classification rates being observed in the bilateral hippocampus. Greater hippocampal activation in children was found for additions as compared to subtractions by De Smedt et al. (2011). Qin et al. (2014) longitudinal study showed that 7-to 9-years-old children showed increases in hippocampus activation and decreases in prefrontal-parietal activation during addition problem solving, suggesting a transition from counting to retrieval (Qin et al., 2014). Rosenberg-Lee et al. (2018) found increases in hippocampus and decreases in fronto-parietal activity when children solved a single-digit addition verification task after they completed an 8-week number and arithmetic training. In summary, studies suggest an important role of the hippocampus for addition, so future studies need to address the role of this brain area in distinguishing between the Schema-based and Fact-retrieval hypotheses.

No Evidence Supporting the Fact-Retrieval Hypothesis in Our Study

The Fact-retrieval hypothesis would have been supported by the finding of greater activation in left MTG/STG at time 1 predicting gains in subtraction fluency or brain activation shifting from parietal to temporal cortex over time. We found no such effects. Our results showed that all children, *regardless of improvement*, showed decreased brain activation in verbal regions over time for large subtractions. The lack of a problem size effect in verbal regions at the first time point and the fact that brain activation in this region decreased over time regardless of improvement argues against the Fact-retrieval hypothesis, suggesting that a shift toward retrieval is not the underlying mechanism for gains in subtraction fluency. Even when looking at the whole brain, no significant differences between improvement groups were found in any region at time 1. For the changes in brain activation over time, greater activation was found for non-improvers as compared to improvers in a region sometimes reported in studies looking at retrieval, the supramarginal gyrus (e.g., Lee, 2000; Rivera et al., 2005). While the exact role of supramarginal gyrus in arithmetic processing is not yet clear, the finding suggests that engaging this region is actually associated with a lack of improvement in fluency. We found no significant brain activation in regions considered to play a role in memory formation, such as the hippocampus (e.g., Cho et al., 2011) or in other regions reported to be activated (or deactivated) when retrieving, such as the angular gyrus (e.g., De Smedt et al., 2011). In line with Thevenot et al. (2016), our study argues against previous evidence suggesting that by the time children are 10-years-old, they rely on retrieval to solve single-digit arithmetic problems (Ashcraft and Fierman, 1982), suggesting instead that different operations recruit distinct neural networks

(Arsalidou and Taylor, 2011; Rosenberg-Lee et al., 2011), even for single-digit problems.

Educational Relevance and Conclusion

Our fMRI study has filled a gap in the literature by providing evidence that early reliance on brain areas implicated in quantity representation is an important predictor explaining gains in subtraction fluency in children. This finding supports the Schema-based hypothesis, and is consistent with previous behavioral (Baroody, 1983; LeFevre et al., 2006; Fayol and Thevenot, 2012) and fMRI (Prado et al., 2011, 2014; Evans et al., 2016) evidence suggesting that children do not rely on retrieval to solve subtractions, but that procedures become more automatic with skill development to support this operation (Fayol and Thevenot, 2012; Barrouillet and Thevenot, 2013). Our study constitutes an example of the utility of neuroimaging to provide important information in order to answer educationally relevant questions.

In our study, we did not give children any instruction in the kind of strategy they should use to solve the task, so it is likely that individuals used different strategies. However, it was the children who relied on quantity mechanisms by engaging parietal cortex early in development the ones who showed greater fluency 2 years later. This finding suggests that the engagement of parietal-based calculation strategies should be encouraged in the classroom to solve subtraction problems. We argue that calculation practice over the course of formal math education will lead to subtractions becoming more automatic. We found no evidence suggesting that the rote memorization of subtraction facts should be encouraged in school.

According to Siegler's adaptive strategy choice model (Siegler and Shipley, 1995), arithmetic strategies are chosen depending on their efficiency. One reason why relying on procedures to solve subtractions might be more efficient than retrieval has to do with their non-commutative nature. While additions and multiplications are commutative so, for example, $3 + 6$ and $6 + 3$ could share common memory nodes (Rickard and Bourne, 1996), subtraction is not. Using retrieval might not be efficient for solving subtractions because children would have to store in memory twice the number of subtraction facts (i.e., $6-3 = 3$, but $3-6 = -3$). As suggested by Campbell and Xue (2001), there should be greater retrieval interference for subtraction facts, making retrieval less efficient for this operation and promoting the use of procedures.

We cannot rule out the possibility that the effects we found in the brain are the consequence of the way subtractions are taught in the United States. Math curriculum in North America emphasizes conceptual understanding over fact mastery (Geary et al., 1996b), with subtractions usually being taught by using counting strategies or inverse addition (Geary et al., 1993; LeFevre et al., 2006), and engaging brain regions involved in finger representations (Berteletti and Booth, 2015a). According to Siegler's distribution of associations model (Siegler and Jenkins, 1989), with experience, certain problems become associated with certain strategies, as do problems with answers. If a problem is consistently associated with a given strategy, then the association between them can be even stronger than the problem-solution association, leading to the application of the most frequently used strategy. Children may also rely more consistently on procedures for subtractions to avoid the switching cost associated with mixing strategies (Lemaire and Lecacheur, 2010). Considering imaging evidence showing that the method of learning arithmetic has a direct impact on the brain (Delazer et al., 2005), it is possible that these teaching differences across countries could play an important role in supporting the Fact-retrieval vs. Schema-based hypotheses. Future studies comparing students from countries having a different emphasis on retrieval should be carried out to test this hypothesis.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Northwestern University Institutional Review Board. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

JB conceptualized and designed the project and supervised the data collection. MS-P and JB formulated the research question. IB contributed to data collection. MS-P analyzed the data and wrote the first draft of the manuscript. All authors contributed to the interpretation of the results, revised the manuscript, and approved the final version for publication.

FUNDING

This research was supported by HD059177 from the National Institute of Child Health and Human Development to JB.

ACKNOWLEDGMENTS

We would like to thank John Binzak and Rachna Mutreja for their help with data collection. We would also like to thank all participating children and their parents.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00163/full#supplementary-material

REFERENCES

Alloway, T. P., Gathercole, S. E., and Pickering, S. J. (2007). *Automated Working Memory Assessment*. London: Pearson Assessment.

Andin, J., Fransson, P., Rönnberg, J., and Rudner, M. (2015). Phonology and arithmetic in the language – calculation network. *Brain Lang.* 143, 97–105. doi: 10.1016/j.bandl.2015.02.004

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nat. Rev. Neurosci.* 9, 278–291. doi: 10.1038/nrn2334

Arsalidou, M., Pawliw-levac, M., Sadeghi, M., and Pascual-leone, J. (2018). Developmental Cognitive Neuroscience Brain areas associated with numbers and calculations in children: meta- analyses of fMRI studies. *Dev. Cogn. Neurosci.* 30, 239–250. doi: 10.1016/j.dcn.2017.08.002

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

Artemenko, C., Soltanlou, M., Ehlis, A. C., Nuerk, H. C., and Dresler, T. (2018). The neural correlates of mental arithmetic in adolescents: a longitudinal fNIRS study. *Behav. Brain Funct.* 14, 1–13. doi: 10.1186/s12993-018-0137-8

Ashcraft, M. H. (1982). The development of mental arithmetic: a chronometric approach. *Dev. Rev.* 2, 213–236. doi: 10.1016/0273-2297(82)90012-0

Ashcraft, M. H., and Fierman, B. A. (1982). Mental addition in third, fourth, and sixth graders. *J. Exp. Child Psychol.* 33, 216–234. doi: 10.1016/0022-0965(82)90017-0

Baroody, A. J. (1983). The development of procedural knowledge: an alternative explanation for chronometric trends of mental arithmetic. *Dev. Rev.* 230, 225–230.

Barrouillet, P., and Lépine, R. (2005). Working memory and children's use of retrieval to solve addition problems. *J. Exp. Child Psychol.* 91, 183–204. doi: 10.1016/j.jecp.2005.03.002

Barrouillet, P., Mignon, M., and Thevenot, C. (2008). Strategies in subtraction problem solving in children. *J. Exp. Child Psychol.* 99, 233–251. doi: 10.1016/j.jecp.2007.12.001

Barrouillet, P., and Thevenot, C. (2013). On the problem-size effect in small additions: Can we really discard any counting-based account? *Cognition* 128, 35–44. doi: 10.1016/j.cognition.2013.02.018

Beauchamp, M., Dagher, A., Aston, J., and Doyon, J. (2003). Dynamic functional changes associated with cognitive skill learning of an adapted. *Neuroimage* 20, 1649–1660. doi: 10.1016/j.neuroimage.2003.07.003

Belfanz, R., Herzog, L., and Mac Iver, D. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: early identification and effective interventions. *Educ. Psychol.* 42, 1–13.

Berteletti, I., and Booth, J. (2015a). "Finger representation and finger-based strategies in the acquisition of number meaning and arithmetic," in *Development of Mathematical Cognition. Neural Substrates and Genetic Influences*, eds D. B. Berch, D. C. Geary, and K. Mann Koepke (San Diego, CA: Academic Press), 109–139.

Berteletti, I., and Booth, J. R. (2015b). Perceiving fingers in single-digit arithmetic problems. *Front. Psychol.* 6:226. doi: 10.3389/fpsyg.2015.00226

Berteletti, I., Man, G., and Booth, J. R. (2014). How number line estimation skills relate to neural activations in single digit subtraction problems. *Neuroimage* 107C, 198–206. doi: 10.1016/j.neuroimage.2014.12.011

Bookheimer, S. (2002). Functional MRI and language: new approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.* 25, 151–188. doi: 10.1146/annurev.neuro.25.112701.142946

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2002). Functional anatomy of intra- and cross-modal lexical tasks. *Neuroimage* 16, 7–22. doi: 10.1006/nimg.2002.1081

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2003). Relation between brain activation and lexical performance. *Hum. Brain Mapp.* 19, 155–169. doi: 10.1002/hbm.10111

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. M. (2004). Development of brain mechanisms for processing orthographic and phonologic representations. *J. Cogn. Neurosci.* 16, 1234–1249. doi: 10.1021/ja064902x.Brilliant

Büchel, C., Coull, J. T., and Friston, K. J. (1999). The predictive value of changes in effective connectivity for human learning. *Science* 283, 1538–1541. doi: 10.1126/science.283.5407.1538

Bush, G., Whalen, P. J., Rosen, B. R., Jenike, M. A., McInerney, S. C., and Rauch, S. L. (1998). The counting stroop: an interference task specialized for functional neuroimaging - Validation study with functional MRI. *Hum. Brain Mapp.* 6, 270–282.

Campbell, J. I. D., and Xue, Q. (2001). Cognitive arithmetic across cultures. *J. Exp. Psychol. Gen.* 130, 299–315. doi: 10.1037//0096-3445.130.2.299

Casey, B. J., Tottenham, N., Liston, C., and Durston, S. (2005). Imaging the developing brain: what have we learned about cognitive development? *Trends Cogn. Sci.* 9, 104–110. doi: 10.1016/j.tics.2005.01.011

Caviola, S., Mammarella, I. C., Pastore, M., and LeFevre, J. A. (2018). Children's strategy choices on complex subtraction problems: individual differences and developmental changes. *Front. Psychol.* 9:1209. doi: 10.3389/fpsyg.2018.01209

Chang, T., Metcalfe, A. W. S., Padmanabhan, A., Chen, T., and Menon, V. (2016). Heterogeneous and nonlinear development of human posterior parietal cortex function. *Neuroimage* 126, 184–195. doi: 10.1016/j.neuroimage.2015.11.053

Chang, T. T., Rosenberg-Lee, M., Metcalfe, A. W. S., Chen, T., and Menon, V. (2015). Development of common neural representations for distinct numerical problems. *Neuropsychologia* 75, 481–495. doi: 10.1016/j.neuropsychologia.2015.07.005

Cho, S., Metcalfe, A. W. S., Young, C. B., Ryali, S., Geary, D. C., and Menon, V. (2012). Hippocampal-prefrontal engagement and dynamic causal interactions in the maturation of children's fact retrieval. *J. Cogn. Neurosci.* 24, 1849–1866. doi: 10.1162/jocn_a_00246

Cho, S., Ryali, S., Geary, D. C., and Menon, V. (2011). How does a child solve 7+8? Decoding brain activity patterns associated with counting and retrieval strategies. *Dev. Sci.* 14, 989–1001. doi: 10.1111/j.1467-7687.2011.01055.x

De Smedt, B., Holloway, I. D., and Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetical fl uency. *Neuroimage* 57, 771–781. doi: 10.1016/j.neuroimage.2010.12.037

Dehaene, S., and Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Math. Cogn.* 1, 83–120. doi: 10.1016/j.neuropsychologia.2006.11.012

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

Delazer, M., Ischebeck, A., Domahs, F., Zamarian, L., Koppelstaetter, F., Siedentopf, C. M., et al. (2005). Learning by strategies and learning by drill — evidence from an fMRI study. *Neuroimage* 25, 838–849. doi: 10.1016/j.neuroimage.2004.12.009

Demir-Lira, Ö. E., Prado, J., and Booth, J. R. (2014). The differential role of verbal and spatial working memory in the neural basis of arithmetic. *Dev. Neuropsychol.* 39, 440–458. doi: 10.1080/87565641.2014.939182

Demir-Lira, Ö. E., Prado, J., and Booth, J. R. (2015). Parental socioeconomic status and the neural basis of arithmetic: differential relations to verbal and visuo-spatial representations. *Dev. Sci.* 5, 1–16. doi: 10.1111/desc.12268

Demir-Lira, Ö. E., Prado, J., and Booth, J. R. (2016). Neural correlates of math gains vary depending on parental socioeconomic status (SES). *Front. Psychol.* 7:892. doi: 10.3389/fpsyg.2016.00892

Eklund, A., Nichols, T. E., and Knutsson, H. (2016). Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7900–7905. doi: 10.1073/pnas.1612033113

Ericsson, K. A., and Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data (Revised Edition)*. Cambridge, MA: MIT Press.

Evans, T. M., Flowers, D. L., Luetje, M. M., Napoliello, E., and Eden, G. F. (2016). Functional neuroanatomy of arithmetic and word reading and its relationship to age. *Neuroimage* 143, 304–315. doi: 10.1016/j.neuroimage.2016.08.048

Fayol, M., and Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction problems. *Cognition* 123, 392–403. doi: 10.1016/j.cognition.2012.02.008

Fedorenko, E., Duncan, J., and Kanwisher, N. (2012). Report language-selective and domain-general regions lie side by side within Broca' s area. *Curr. Biol.* 22, 2059–2062. doi: 10.1016/j.cub.2012.09.011

Fedorenko, E., Hsieh, P. J., Nieto-Castañón, A., Whitfield-Gabrieli, S., and Kanwisher, N. (2010). New method for fMRI investigations of language:

defining ROIs functionally in individual subjects. *J. Neurophysiol.* 104, 1177–1194. doi: 10.1152/jn.00032.2010

Fedorenko, E., and Kanwisher, N. (2009). Neuroimaging of language: why hasn't a clearer picture emerged? *Linguist. Lang. Compass* 3, 839–865. doi: 10.1111/j.1749-818X.2009.00143.x

Garavan, H., Kelley, D., Rosen, A., Rao, S. M., and Stein, E. A. (2000). Practice-related functional activation changes in a working memory task. *Microsc. Res. Tech.* 51, 54–63. doi: 10.1002/1097-0029(20001001)51:1<54::AID-JEMT6>3.0.CO;2-J

Geary, D. C. (1994). *Children's Mathematical Development: Research and Practical Applications*. Washington, DC: American Psychological Association.

Geary, D. C., Bow-Thomas, C. C., Liu, F., and Siegler, R. S. (1996a). Development of arithmetical competencies in Chinese and American Children: influence of age, language, and schooling. *Child Dev.* 67, 2022–2044. doi: 10.1111/j.1467-8624.1996.tb01841.x

Geary, D. C., Chen, G. P., Salthouse, T. A., and Fan, L. (1996b). Are East Asian versus American differences in arithmetical ability a recent phenomenon? *Dev. Psychol.* 32, 254–262. doi: 10.1037/0012-1649.32.2.254

Geary, D. C., Frensch, P. A., and Wiley, J. G. (1993). Simple and complex mental subtraction: strategy choice and speed-of-processing differences in younger and older adults. *Psychol. Aging* 8, 242–256. doi: 10.1037/0882-7974.8.2.242

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., and Neuper, C. (2009). To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. *Neuropsychologia* 47, 604–608. doi: 10.1016/j.neuropsychologia.2008.10.013

Grabner, R. H., Brunner, C., Lorenz, V., Vogel, S. E., and De Smedt, B. (2020). Fact retrieval or compacted procedures in arithmetic – a neurophysiological investigation of two hypotheses. *bioRxiv* [Preprint]. doi: 10.1101/2020.03.10.985143

Hempel, A., Giesel, F. L., Garcia Caraballo, N. M., Amann, M., Meyer, H., Wüstenberg, T., et al. (2004). Plasticity of cortical activation related to working memory during training. *Am. J. Psychiatry* 161, 745–747. doi: 10.1176/appi.ajp.161.4.745

Ischebeck, A., Zamarian, L., Siedentopf, C., Koppelsta, F., Benke, T., Felber, S., et al. (2006). How specifically do we learn? Imaging the learning of multiplication and subtraction. *Neuroimage* 30, 1365–1375. doi: 10.1016/j.neuroimage.2005.11.016

Iuculano, T., Rosenberg-Lee, M., Richardson, J., Tenison, C., Fuchs, L., Supekar, K., et al. (2015). Cognitive tutoring induces widespread neuroplasticity and remediates brain function in children with mathematical learning disabilities. *Nat. Commun.* 6:8453. doi: 10.1038/ncomms9453

Karmiloff-Smith, A. (2010). "Neuroimaging of the developing brain: taking "developing" seriously," in *Human Brain Mapping*, Vol. 36 (London: University of London), 934–941. https://doi.org/10.1002/hbm.21074

Kawashima, R., Taira, M., Okita, K., Inoue, K., and Tajima, N. (2004). A functional MRI study of simple arithmetic — a comparison between children and adults. 18, 225–231. doi: 10.1016/j.cogbrainres.2003.10.009

Kirk, E. P., and Ashcraft, M. H. (2001). Telling stories: the perils and promise of using verbal reports to study math strategies. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 157–175. doi: 10.1037/0278-7393.27.1.157

Knops, A., Thirion, B., Hubbard, E. M., Michel, V., and Dehaene, S. (2009). Recruitment of an area involved in eye movements during mental arithmetic. *Science* 324, 1583–1586. doi: 10.1126/science.1171599

Krings, T., Töpper, R., Foltys, H., Erberich, S., Sparing, R., Willmes, K., et al. (2000). Cortical activation patterns during complex motor tasks in piano players and control subjects. A functional magnetic resonance imaging study. *Neurosci. Lett.* 278, 189–193. doi: 10.1016/S0304-3940(99)00930-1

Krinzinger, H. (2011). The role of finger representations and saccades for number processing: an fMRI study in children. *Front. Psychol.* 2:373. doi: 10.3389/fpsyg.2011.00373

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: 10.1080/87565640802101474

Lee, K. M. (2000). Cortical areas differentially involved in multiplication and subtraction: a functional magnetic resonance imaging study and correlation with a case of selective acalculia. *Ann. Neurol.* 48, 657–661.

Lefevre, J., Sadesky, G. S., and Bisanz, J. (1996). Selection of procedures in mental addition: reassessing the problem size effect in adults. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 216–230.

LeFevre, J. A., DeStefano, D., Penner-Wilger, M., and Daley, K. E. (2006). Selection of procedures in mental subtraction. *Can. J. Exp. Psychol.* 60, 209–220. doi: 10.1037/cjep2006020

Lemaire, P., and Lecacheur, M. (2010). Strategy switch costs in arithmetic problem solving. *Mem. Cogn.* 38, 322–332. doi: 10.3758/MC.38.3.322

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., and Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19, 1233–1239. doi: 10.1016/S1053-8119(03)00169-1

Matejko, A. A., and Ansari, D. (2016). Trajectories of symbolic and nonsymbolic magnitude processing in the first year of formal schooling. *PLoS One* 11:149863. doi: 10.1371/journal.pone.0149863

Mathieu's, R., Epinat-Duclos, J., Sigovan, M., Breton, A., Cheylus, A., Fayol, M., et al. (2018). What's Behind a "+" Sign? Perceiving an arithmetic operator recruits brain circuits for spatial orienting. *Cereb. Cortex* 28, 1673–1684. doi: 10.1093/cercor/bhx064

Mathieu's, R., Gourjon, A., Couderc, A., Thevenot, C., and Prado, J. (2016). Running the number line: rapid shifts of attention in single-digit arithmetic. *Cognition* 146, 229–239. doi: 10.1016/j.cognition.2015.10.002

Mazaika, P., Hoeft, F., Glover, G., and Reiss, A. (2009). "Methods and software for fMRI analysis for clinical subjects," in *Proceeding of the 15th Annual Meeting The Organization of Human Brain Mapping, San Francisco, CA*, (Minneapolis, MN: The Organization of Human Brain Mapping).

Mazaika, P. K., Whitfield-Gabrieli, S., and Reiss, A. L. (2007). Artifact repair of fMRI data from high motion clinical subjects. *Neuroimage* 36(Suppl. 1):S142.

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

Mussolin, C., Nys, J., Content, A., and Leybaert, J. (2014). Symbolic number abilities predict later approximate number system acuity in preschool children. *PLoS One* 9:e91839. doi: 10.1371/journal.pone.0091839

Nieto-Castañón, A., and Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *Neuroimage* 63, 1646–1669. doi: 10.1016/j.neuroimage.2012.06.065

Núñez-Peña, M. I., Colomé, A., and Tubau, E. (2015). Efficiency of arithmetic procedures modulates the problem-size effect in subtraction. *Span. J. Psychol.* 18:E41. doi: 10.1017/sjp.2015.43

Prat, C. S., and Just, M. A. (2011). Exploring the neural dynamics underpinning individual differences in sentence comprehension. *Cereb. Cortex* 21, 1747–1760. doi: 10.1093/cercor/bhq241

Prat, C. S., Keller, T. A., and Just, M. A. (2007). Individual differences in sentence comprehension: a functional magnetic resonance imaging investigation of syntactic and lexical processing demands. *J. Cogn. Neurosci.* 19, 1950–1963. doi: 10.1162/jocn.2007.19.12.1950

Peters, G., de Smedt, B., Torbeyns, J., Ghesquière, P., and Verschaffel, L. (2012). Children's use of subtraction by addition on large single-digit subtractions. *Educ. Stud. Math.* 79, 335–349. doi: 10.1007/s10649-011-9308-3

Piazza, M., Pinel, P., Le Bihan, D., Dehaene, S., and Cedex, O. (2004). Tuning curves for approximate numerosity in the human Intraparietal Sulcus. *Neuron* 44, 547–555. doi: 10.1016/j.neuron.2004.10.014

Pinel, P., Dehaene, S., Rivière, 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., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., and Gabrieli, J. D. E. E. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage* 35, 15–35. doi: 10.1006/nimg.1999.0441

Pollack, C., and Ashby, N. C. (2017). Where arithmetic and phonology meet: the meta-analytic convergence of arithmetic and phonological processing in the brain. *Dev. Cogn. Neurosci.* 30, 251–264. doi: 10.1016/j.dcn.2017.05.003

Polspoel, B., Peters, L., Vandermosten, M., and De Smedt, B. (2017). Strategy over operation: neural activation in subtraction and multiplication during fact retrieval and procedural strategy use in children. *Hum. Brain Mapp.* 38, 4657–4670. doi: 10.1002/hbm.23691

Prado, J., Mutreja, R., and Booth, J. R. (2014). Developmental dissociation in the neural responses to simple multiplication and subtraction problems. *Dev. Sci.* 17, 537–552. doi: 10.1111/desc.12140

Prado, J., Mutreja, R., Zhang, H., Mehta, R., Desroches, A. S., Minas, J. E., et al. (2011). Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. *Hum. Brain Mapp.* 32, 1932–1947. doi: 10.1002/hbm.21159

Price, G. R., Mazzocco, M. M. M., and Ansari, D. (2013). Why mental arithmetic counts: brain activation during single digit arithmetic predicts high school math scores. *J. Neuroci.* 33, 156–163. doi: 10.1523/JNEUROSCI.2936-12.2013

Price, G. R., Wilkey, E. D., Yeo, D. J., and Cutting, L. E. (2016). The relation between 1st grade grey matter volume and 2nd grade math competence. *Neuroimage* 124, 232–237. doi: 10.1016/j.neuroimage.2015.08.046

Qin, S., Cho, S., Chen, T., Rosenberg-Lee, M., Geary, D. C., and Menon, V. (2014). Hippocampal-neocortical functional reorganization underlies children's cognitive development. *Nat. Neurosci.* 17, 1263–1269. doi: 10.1038/nn.3788

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., Romero, S. G., Basso, G., Wharton, C., Flitman, S., and Grafman, J. (2000). The calculating brain: an fMRI study. *Neuropsychologia* 38, 325–335. doi: 10.1016/s0028-3932(99)00068-8

Ritchie, S. J., and Bates, T. C. (2013). Enduring links from childhood mathematics and reading achievement to adult socioeconomic status. *Psychol. Sci.* 24, 1301–1308. doi: 10.1177/0956797612466268

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., Ashkenazi, S., Chen, T., Young, C. B., Geary, D. C., and Menon, V. (2015). Brain hyper-connectivity and operation-specific deficits during arithmetic problem solving in children with developmental dyscalculia. *Dev. Sci.* 18, 351–372. doi: 10.1111/desc.12216

Rosenberg-Lee, M., Chang, T. T., Young, C. B., Wu, S., and Menon, V. (2011). Functional dissociations between four basic arithmetic operations in the human posterior parietal cortex: a cytoarchitectonic mapping study. *Neuropsychologia* 49, 2592–2608. doi: 10.1016/j.neuropsychologia.2011.04.035

Rosenberg-Lee, M., Iuculano, T., Bae, S. R., Richardson, J., Qin, S., Jolles, D., et al. (2018). Short-term cognitive training recapitulates hippocampal functional changes associated with one year of longitudinal skill development. *Trends Neurosci. Educ.* 10, 19–29. doi: 10.1016/j.tine.2017.12.001

Rypma, B., and D'Esposito, M. (1999). The roles of prefrontal brain regions in components of working memory: effects of memory load and individual differences. *Proc. Natl. Acad. Sci. U.S.A.* 96, 6558–6563. doi: 10.1073/pnas.96.11.6558

Seyler, D. J., Kirk, E. P., and Ashcraft, M. H. (2003). Elementary subtraction. *J. Exp. Psychol. Learn. Mem. Cogn.* 29, 1339–1352. doi: 10.1037/0278-7393.29.6.1339

Siegler, R. S. (1987). "Strategy choice in subtraction," in *Cognitive Processes in Mathematics*, eds J. A. Sloboda, and D. Rogers (New York, NY: Oxford University), 81–106.

Siegler, R. S. (1989). Hazards of mental chronometry: an example from children's subtraction. *J. Educ. Psychol.* 81, 497–506. doi: 10.1037/0022-0663.81.4.497

Siegler, R. S., and Jenkins, E. (1989). *How Children Discover New Strategies*. Hillsdale, NJ: Erlbaum.

Siegler, R. S., and Shipley, C. (1995). "Variation, selection, and cognitive change," in *Developing Cognitive Competence: New Approaches to Process Modelling*, eds T. Simon, and G. Halford (Hillsdale, NJ: Erlbaum), 31–76.

Simon, O., Mangin, J. F., Cohen, L., Le Bihan, D., and Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron* 33, 475–487. doi: 10.1016/S0896-6273(02)00575-5

Sokolowski, H. M., Fias, W., Mousa, A., and Ansari, D. (2017). Common and distinct brain regions in both parietal and frontal cortex support symbolic and nonsymbolic number processing in humans: a functional neuroimaging meta-analysis. *Neuroimage* 146, 376–394. doi: 10.1016/j.neuroimage.2016.10.028

Suárez-Pellicioni, M., and Booth, J. R. (2018). Fluency in symbolic arithmetic refines the approximate number system in parietal cortex. *Hum. Brain Mapp.* 39, 3956–3971. doi: 10.1002/hbm.24223

Suárez-Pellicioni, M., Fuchs, L., and Booth, J. R. (2019a). Temporo-frontal activation during phonological processing predicts gains in arithmetic facts in young children. *Dev. Cogn. Neurosci.* 40:100735. doi: 10.1016/j.dcn.2019.100735

Suárez-Pellicioni, M., Lytle, M., Younger, J. W., and Booth, J. R. (2019b). A longitudinal neuroimaging dataset on arithmetic processing in school children. *Sci. Data* 6:190040. doi: 10.1038/sdata.2019.40

Suárez-Pellicioni, M., Prado, J., and Booth, J. (2018). Lack of improvement in multiplication is associated with reverting from verbal retrieval to numerical operations. *Neuroimage* 183, 859–871. doi: 10.1016/j.neuroimage.2018.08.074

Thevenot, C., Barrouillet, P., Castel, C., and Uittenhove, K. (2016). Ten-year-old children strategies in mental addition: a counting model account. *Cognition* 146, 48–57. doi: 10.1016/j.cognition.2015.09.003

Uittenhove, K., Thevenot, C., and Barrouillet, P. (2016). Fast automated counting procedures in addition problem solving: When are they used and why are they mistaken for retrieval? *Cognition* 146, 289–303. doi: 10.1016/j.cognition.2015.10.008

Welcome, S. E., and Joanisse, M. F. (2012). Individual differences in skilled adult readers reveal dissociable patterns of neural activity associated with component processes of reading. *Brain Lang.* 120, 360–371. doi: 10.1016/j.bandl.2011.12.011

Weschler, D. (1999). *Weschler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.

Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Itasca, IL: Riverside Publishing.

Woods, S. S., Resnick, L. B., and Groen, G. J. (1975). An experimental test of five process models for subtraction. *J. Educ. Psychol.* 67, 17–21. doi: 10.1037/h0078666

Conflict of Interest: 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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 July 2020
doi: 10.3389/fpsyg.2020.01115

Efficacy of a Computer-Based Learning Program in Children With Developmental Dyscalculia. What Influences Individual Responsiveness?

Juliane Kohn[1,2], Larissa Rauscher[3], Karin Kucian[4,5], Tanja Käser[6], Anne Wyschkon[2], Günter Esser[1,2] and Michael von Aster[1,4,7]*

[1] Department of Psychology, University of Potsdam, Potsdam, Germany, [2] Academy of Psychotherapy and Intervention Research, University of Potsdam, Potsdam, Germany, [3] Department of Child and Adolescent Psychiatry, German Red Cross Hospital, Berlin, Germany, [4] Center for MR Research, University Children's Hospital Zürich, Zurich, Switzerland, [5] Children's Research Center, University Children's Hospital Zürich, Zurich, Switzerland, [6] Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, [7] Center of School and Mental Rehabilitation, German Red Cross Hospitals, Berlin, Germany

OPEN ACCESS

Edited by:
Korbinian Moeller,
Loughborough University,
United Kingdom

Reviewed by:
Matthias Hartmann,
University of Bern, Switzerland
Evelyn Kroesbergen,
Radboud University Nijmegen,
Netherlands

***Correspondence:**
Juliane Kohn
jkohn@uni-potsdam.de

Specialty section:
This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 15 October 2019
Accepted: 30 April 2020
Published: 15 July 2020

Citation:
Kohn J, Rauscher L, Kucian K, Käser T, Wyschkon A, Esser G and von Aster M (2020) Efficacy of a Computer-Based Learning Program in Children With Developmental Dyscalculia. What Influences Individual Responsiveness? Front. Psychol. 11:1115. doi: 10.3389/fpsyg.2020.01115

This study presents the evaluation of a computer-based learning program for children with developmental dyscalculia and focuses on factors affecting individual responsiveness. The adaptive training program Calcularis 2.0 has been developed according to current neuro-cognitive theory of numerical cognition. It aims to automatize number representations, supports the formation and access to the mental number line and trains arithmetic operations as well as arithmetic fact knowledge in expanding number ranges. Sixty-seven children with developmental dyscalculia from second to fifth grade (mean age 8.96 years) were randomly assigned to one of two groups (Calcularis group, waiting control group). Training duration comprised a minimum of 42 training sessions à 20 min within a maximum period of 13 weeks. Compared to the waiting control group, children of the Calcularis group demonstrated a higher benefit in arithmetic operations and number line estimation. These improvements were shown to be stable after a 3-months post training interval. In addition, this study examines which predictors accounted for training improvements. Results indicate that this self-directed training was especially beneficial for children with low math anxiety scores and without an additional reading and/or spelling disorder. In conclusion, Calcularis 2.0 supports children with developmental dyscalculia to improve their arithmetical abilities and their mental number line representation. However, it is relevant to further adapt the setting to the individual circumstances.

Keywords: developmental dyscalculia, mathematics instruction, computer-based training, intelligent tutoring system (ITS), numerical development, evaluative study, primary school

INTRODUCTION

Solid mathematic skills are not only important for a child's academic career but are also necessary for numerous situations in every-day life. A weakness in this area cannot only lead to school-related problems but may also affect occupational routes and emotional well-being (Cohen Kadosh et al., 2013). Children with developmental dyscalculia (DD) demonstrate highly diverse performance profiles (Kaufmann and von Aster, 2012) with deficits regarding basic numerical processing, transcoding, counting, arithmetic fact retrieval, basic arithmetic skills, and word problems (e.g., Geary et al., 2007; Kaufmann et al., 2013; Kuhn et al., 2013; Landerl, 2013). Due to different definition and diagnostic criteria, the prevalence of DD in English and German speaking children vary between 1.8 and 5% (Lewis et al., 1994; Esser et al., 2008; Fischbach et al., 2013).

Several studies have demonstrated that targeted interventions can improve different aspects of numerical cognition in children with DD (Dowker, 2004; Bryant et al., 2008; Fuchs et al., 2010). Ise et al. (2012) conducted a meta-analysis concerning the efficacy of different treatment approaches for children with mathematical disabilities and reported a moderate mean effect size (Hedges' $g = 0.50$) which is comparable to the results of other meta-analyses (Baker et al., 2002; Kroesbergen and van Luit, 2003; Chodura et al., 2015).

Based on empirical evidence different characteristics of effective treatments of children with DD are proposed in the literature. Treatment approaches are considered to be especially effective, when they are adaptive to the child's learning needs and learning speed (Burns et al., 2010; Moeller et al., 2012). Children with DD benefit from a structured design, hierarchical organization and frequent as well as constant repetition and practice (Fuchs et al., 2008). Reward systems enhance the children's motivation to solve arithmetic problems (Fuchs et al., 2008; Butterworth and Laurillard, 2010). Since children with DD show diverse deficits, effective training approaches need to address multiple areas of numerical cognition such as basic numerical competencies, conceptual and procedural knowledge and arithmetic fact retrieval (Kaufmann et al., 2003; Dowker, 2007).

During the last years several computer-assisted training systems have been developed.

Those training programs do not aim to replace classic learning therapy interventions conducted by therapists or special need teachers but aim to support the development and automatization of specific cognitive components in the numerical domain (von Aster et al., 2012). In particular, for children with DD a computerized training to enhance numerical cognition offers considerable advantages (Räsänen et al., 2015, 2019). It allows addressing an optimal level of difficulty and learning speed through an individually customized task selection. So called intelligent tutoring systems (ITS) are able to build up an internal image of the learner's skill and ability profile in form of a "learner model" by studying the child's actions (von Aster and Lipka, 2018).

Furthermore, a computerized training offers the possibility of immediate feedback about the correctness of a solved task. Direct chronological proximity is central for knowledge acquisition (Krajewski and Ennemoser, 2010). To support this, adaptive computer-based trainings can introduce tasks being slightly challenging and thus may foster the development of new skills. Additionally, the computer represents an attractive learning medium (Kulik and Kulik, 1991; Schoppek and Tullis, 2010) providing intensive training in a stimulating environment (Kullik, 2004). Particularly for children with DD a computerized training provides the possibility of a learning environment detached from competitive performance pressure and peer comparisons in the classroom context and offers a less stressful and socially risk-free setting to explore mathematics (Käser and von Aster, 2013). This is especially important, since the repeated experience of failure may lead to math anxiety or negative attitudes toward the subject or the teacher, which in turn may decrease the achievement potential and learning ability (Ashcraft and Faust, 1994; Kohn et al., 2013).

An overview of different computer-assisted interventions can be found in Räsänen et al. (2015, 2019). Interventions can be differentiated according to their content: training of basic numerical competencies like magnitude comparison, mental number line, or subitizing (e.g., *Number Race* – Wilson et al., 2006; Räsänen et al., 2009; *Rescue Calcularis* – Fischer et al., 2008; Kucian et al., 2011), training of arithmetic fact knowledge (Fuchs et al., 2006) or training of a combination of basic-numerical skills, spatial number representation and (simple) arithmetic facts (Butterworth and Laurillard, 2010; Butterworth et al., 2011; *Calcularis* – Käser et al., 2013a; *Meister Cody* – Kuhn and Holling, 2014).

Different meta-analyses examined the effects of computer-based mathematic instruction, revealing positive effects. For example, Li and Ma (2010) reported an average effect size of 0.28 for computer-based math instruction. They found larger effects for elementary school than for higher education and showed that especially children with learning disabilities benefit from computer-based instruction. Other meta-analyses reported positive (immediate) effects with effect sizes ranging from 0.13 to 0.80 (Kulik, 1994; Fletcher-Flinn and Gravatt, 1995; Kroesbergen and van Luit, 2003; Slavin and Lake, 2008; Ise et al., 2012; Chodura et al., 2015). Only very few studies report additional results concerning long-term effects of computer-based training programs (Chodura et al., 2015). According to recent research (meta-analysis) in secondary schools, training programs with high adaptivity to the individual needs of the user outperformed less adaptive types of tutoring systems (Hillmayr et al., 2017).

Additionally, meta-analyses emphasize that the evaluative studies vary highly with respect to sample size, inclusion criteria (severity of math disorders) and outcome variables which influence quality of research and comparability (Seo and Bryant, 2009; Ise et al., 2012; Chodura et al., 2015). A meta-analysis focusing on interventions for children with math difficulties (Chodura et al., 2015) indicated that in at least half of the identified studies a less stringent criterion than recommended by DSM-5 was used to select the study participants, e.g., a rank below the 26th percentile in a standardized mathematical test.

One important step to gain knowledge about the efficacy of training is to understand which circumstances render computer-based training successful and which factors predict training induced improvement (Räsänen, 2015).

So far only few studies addressed this question. For example, Nemmi et al. (2016) found differentiated effects of a combination of a computer-based number line training (NLT) and a computer-based number working memory training (WMT) for children who differ in working memory capacities as well as in mathematic skills. The authors used four training conditions (NLT/reading, WMT/reading, and NLT/WM and reading). While overall the combined training was most effective, they found significant interactions with baseline scores. For example, children with higher working memory capacity reached higher gains (mathematical ability) through the working memory training compared to the number line training. On the other hand, children with higher math performance at baseline benefited more from the number line training.

Another potential predictor for training induced improvement is the coexistence of a reading/spelling disorder. Powell et al. (2009) analyzed differential effects of tutoring (partly computer-assisted instruction) for third-grade students with math difficulties and with or without co-occurring reading difficulties. The study demonstrated a better responsiveness to fact retrieval tutoring on fact retrieval skills for children without a co-occurring reading disorder. In fact, children with a combined disorder did not benefit from the fact retrieval intervention compared to a no treatment condition. It is assumed that children with co-occurring math and reading disabilities show underlying phonological processing deficits (Hecht et al., 2001; Robinson et al., 2002). Therefore, these children could have more severe or various problems performing arithmetic procedures (e.g., counting strategies) as well as retrieving arithmetic facts (e.g., von Aster, 1994, 2000; Geary et al., 2000). Furthermore, results of studies analyzing differences in working memory indicate that the children with double deficits are outperformed by children exhibiting a math disorder in verbal and visuospatial tasks (*Meta-analysis*, Swanson et al., 2009).

One significant non-cognitive factor influencing math performance that received much attention during the last years is math anxiety. Math anxiety is defined as a negative emotional reaction that is characterized by feelings of tension, apprehension, or even dread that interferes with the manipulation of numbers and the solving of mathematical problems in a wide variety of ordinary life and academic situations (cf. Richardson and Suinn, 1972, p. 551; Ashcraft and Faust, 1994, p. 98).

Previous studies have shown that math anxiety can have an adverse effect on longer-term career choices and professional success (Hembree, 1990; Meece et al., 1990; Ma, 1999). In recent years, there are several studies that illustrate a negative relationship between math anxiety and math performance in early elementary school (Wu et al., 2012; Kohn et al., 2013; Ramirez et al., 2013; Vukovic et al., 2013). It is assumed that math anxious students tend to avoid math-related tasks and situations (Ashcraft and Faust, 1994; Ashcraft et al., 2007). They show less confrontation with mathematic tasks, learn less and as a consequence show reduced achievement scores. In addition, they probably receive more negative feedback which increases in turn math anxiety, contributing to a vicious circle (Krinzinger and Kaufmann, 2006; Dowker et al., 2012; von Aster et al., 2017). Furthermore, it is postulated that math anxiety works as a dual task during task processing that reduces working memory capacity which worsens task performance (Ashcraft and Kirk, 2001; Ashcraft et al., 2007; Ashcraft and Moore, 2009). These assumptions regarding ways of explaining the link between math anxiety and mathematics performance are integrated in the Reciprocal Theory (Carey et al., 2016) that postulates a bidirectional relationship. Supekar et al. (2015) found a significant reduction of math anxiety in students with high math anxiety scores at baseline using a one-to-one math tutoring approach. Beyond these behavioral performance effects, they even report that the brain activity levels in the amygdala of high anxious third-grade children normalizes after the intervention to the level of their peers without math anxiety. Concerning math achievement, both groups (high and low anxious children, grade 3) improved their performance in an arithmetic problem solving task equally, as there was no interaction with math anxiety level. Recent work by Kucian et al. (2018a) has demonstrated that math anxiety is even related to changes in brain structure. Particularly, the volume of the amygdala was reduced, which represents the key area in our brain for negative emotional processing such as fear, stress and anxiety. This growing knowledge underscores the important role of emotional factors in mathematical cognition and emphasizes the far-reaching outcome math anxiety can have.

In summary, there are computer-based programs which have been shown to be effective in enhancing number processing, but most of the available programs provide only limited individual adaptability.

Furthermore, evaluative studies rarely use strict criteria for identification of dyscalculic children (Chodura et al., 2015) and lack to investigate long-term effects. In addition, there are only a few studies that focus on individual differences in response to computer-based math instruction and to our knowledge there seems to be none that addresses dyscalculic children.

Based on the need for research for long-term effects of training effects as well as individual responsiveness in dyscalculic children, the objective of the present study is to evaluate the efficacy of the computer-based training program Calcularis 2.0.

Calcularis 2.0 is based on theoretical neurocognitive foundations of numerical cognition, such as the triple-code model (Dehaene, 1992), the four-step developmental model (von Aster and Shalev, 2007) and further theoretical advancements (see i.e., Kucian and Kaufmann, 2009). In particular, we postulate the existence of a core cognitive magnitude system, which enables even different animal species and also human newborns to discriminate large from small numerosities [that are represented from the right (large) to the left (small) space; Kucian et al., 2018b; Di Giorgio et al., 2019], onto which - in the human neuro-cognitive development – non-symbolic numerical meanings are successively and hierarchically transformed into different symbolic number representations (linguistic number word system, visual Arabic notational system, and spatially oriented mental number line). These growing domain-specific cognitive number representations become neuronally built in different and interconnected areas of the brain and act as tools for learning and performing mental arithmetic and higher mathematical reasoning. They are developmentally dependent on environmentally nurtured sensory-motor and cognitive experiences in the pre- and primary school years, especially on increasing capacities of domain-general cognitive abilities like visual-spatial processing, language, working memory and attentional span.

This process of domain-specific representational transformation, which develops from the early non-symbolic perceptions of numerical magnitude, across the acquisition of culturally transmitted symbolization systems (linguistic, visual Arabic) to a gradually expanding, spatially organized symbolic mental number line may be framed by concepts of general cognitive development like the theory of 'Representational Redescription (RR)' postulated by Karmiloff-Smith (1992). RR defines domain-specific cognitive development as (i) being initially constrained by innate predispositions, and (ii) being developmentally formed by the child's experiences in the physical and social environment, in which early implicit procedural representations are successively redescribed into higher order explicit declarative representations, that are mediated by the domain-general information processing system. Importantly, the RR model has been validated empirically in a large number of studies with typically and atypically developing children, including those with Williams-Syndrome and autism spectrum disorder (Karmiloff-Smith, 1998).

From this theoretical point of view the complex development of number processing and calculation abilities may be disturbed or interrupted at different levels of development and for different etiological reasons relating to different dysfunctional components. Hence, it is not surprising that DD is characterized by highly variable clinical pictures including various possible comorbid conditions (Kaufmann et al., 2011). Therefore, intervention strategies should be highly adaptive to individual demands. Furthermore, they should focus on establishing and automatizing the main representational formats of number magnitudes, including the related transcoding routines, while gradually learning and automatizing arithmetic procedures and fact knowledge. Calcularis 2.0 was developed based on these theoretical assumptions and offers children with DD an approach to deal with different deficits. Calcularis 2.0 is a highly adaptive computer-based training program that combines basic numerical cognition with different number representations and arithmetic abilities.

The present evaluation includes a large sample size of children with DD (using strict criteria for identification). Participants were randomly assigned to the Calcularis group completing a 12-weeks training or to the (waiting) control group receiving no training.

We hypothesize that the Calcularis group shows immediate training effects with medium effect sizes, i.e., demonstrate an increased level of arithmetic performance, basic number processing and spatial number representation compared to the (waiting) control group.

We further predict that there is no stronger increase in performance in domains that were not trained (reading, spelling) compared to the control group, indicating domain specificity of the training. Furthermore, we assess the stability of the training effects after a 3-months interval. We hypothesize that there is an increase or at least a consistent level of performance within the Calcularis group. In addition, we examine the impact of different baseline factors on the individual response to the training. As potentially influencing factors we postulate math anxiety, intellectual ability and the coexistence of a reading/spelling disorder. Specifically, we expect that higher improvement goes along with lower math anxiety scores because we assume that math anxiety could work as an impairing factor for deep engagement with the training content (Ashcraft et al., 2007). Additionally, we assume that children with DD and higher intellectual ability have the potential to reach higher gains (Nemmi et al., 2016) and that children without an additional reading and/or spelling disorder tend to show higher profits (Powell et al., 2009).

MATERIALS AND METHODS

Introduction to Calcularis 2.0

Calcularis 2.0 (von Aster et al., 2016) is a highly adaptive computer-based training program. The program's theoretical neurocognitive foundation of numerical cognition and development consists of the triple-code model (Dehaene, 1992), the four-step developmental model (von Aster and Shalev, 2007) and further theoretical advancements (Kucian and Kaufmann, 2009).

The program aims to automatize the different number representations, to support the formation and access to the mental number line and to train arithmetic operations as well as arithmetic fact knowledge in expanding number ranges from 0–10 until 0–1,000.

Calcularis consists of different instructional games, which are hierarchically structured according to number ranges and can be further divided into two areas. The first area focuses on different number representations as well as number processing in general. Transcoding between alternative representations (based on triple code model, Dehaene, 1992) is trained and children learn the three principles of number understanding: cardinality, ordinality, and relativity. Games in this area are hierarchically ordered according to the four-step developmental model (von Aster and Shalev, 2007).

The second area covers cognitive operations and procedures with numbers. In this area, children learn the concepts of arithmetic operations and automate them. The difficulty of the tasks is determined by the complexity of the task, the magnitude of numbers involved and the visual aids available to solve the task. In both areas, games can be categorized based on their complexity. 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.

A consistent number notation that accentuates the properties of numbers is used throughout the training program. The notation is encoded by color, form and topology.

Calcularis 2.0 features a user model allowing flexible adaptation based on the internally mapped learning and knowledge profile of the individual child. The mathematical knowledge trained in the game is divided into more than 250 different fine-grained skills [e.g., "writing a (verbally) given number between 0 and 100," "estimating the quantity of a set of dots," and "adding to numbers between 0 and 10"]. The skills are hierarchically ordered in a directed acyclic graph called dynamic Bayesian network. Connections between the different skills indicate their relations, i.e., it is for example assumed that being able to add two numbers between 0 and 10 is a prerequisite for adding two numbers between 0 and 100. Each skill is associated with a game. When the child plays the associated game of a skill, the system infers from the correct or wrong answers of the child, how well the child already knows this skill. Since the skills are connected, the system at the same time gains also information about the child's knowledge of other skills. The representation of the skills as a graph has another big advantage: every child can follow its individual learning path through the network. Some kids will follow the most direct path through the network, training only a subset of the skills. Other kids will have to backtrack and extensively cover the skills in the area they have deficits. Additionally, an error library with typical error patterns allows to provide targeted games for the remediation of specific mistakes. The high adaptivity differentiates Calcularis 2.0 from other computer-based intervention programs that mostly provide only limited adaptability by means of adapting the task difficulty.

Calcularis 2.0 represents an extended and modified version of Calcularis (Käser et al., 2012, 2013b). The new version includes additional games to train number and quantity comparisons, subitizing (structured and non-structured stimuli), addition and subtraction based on "concrete" material and multiplication and division (**Figure 1**, top). Additionally, the number range 0–20 is explicitly modeled. The program includes an interactive avatar guiding the child through the training and explaining the games. Additionally, a reward system (a virtual zoo) reacting to the individual child's learning progress was implemented to increase the child's motivation and enhance the enjoyment in learning. The virtual zoo allows children to buy and feed animals which can be assigned to various zoo worlds (**Figure 1**, bottom). Calcularis, the pre-version of Calcularis 2.0, was evaluated in children with mathematical difficulties as well as in normally achieving children (Käser et al., 2013a; Rauscher et al., 2016). The study results demonstrated that children benefited significantly from the training regarding spatial number representation and subtraction.

Study Design and Sample

Participants were classified as having DD based on the Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-5) of the American Psychological Association (American Psychiatric Association, 2013).

Criteria for DD were met if a child's performance in a standardized mathematics test (Rechenfertigkeiten- und Zahlenverarbeitungs-Diagnostikum for the 2nd to 6th grade, RZD 2–6, Jacobs and Petermann, 2005) was 1.5 standard deviations ($T \leq 35$) below the average in the speed or

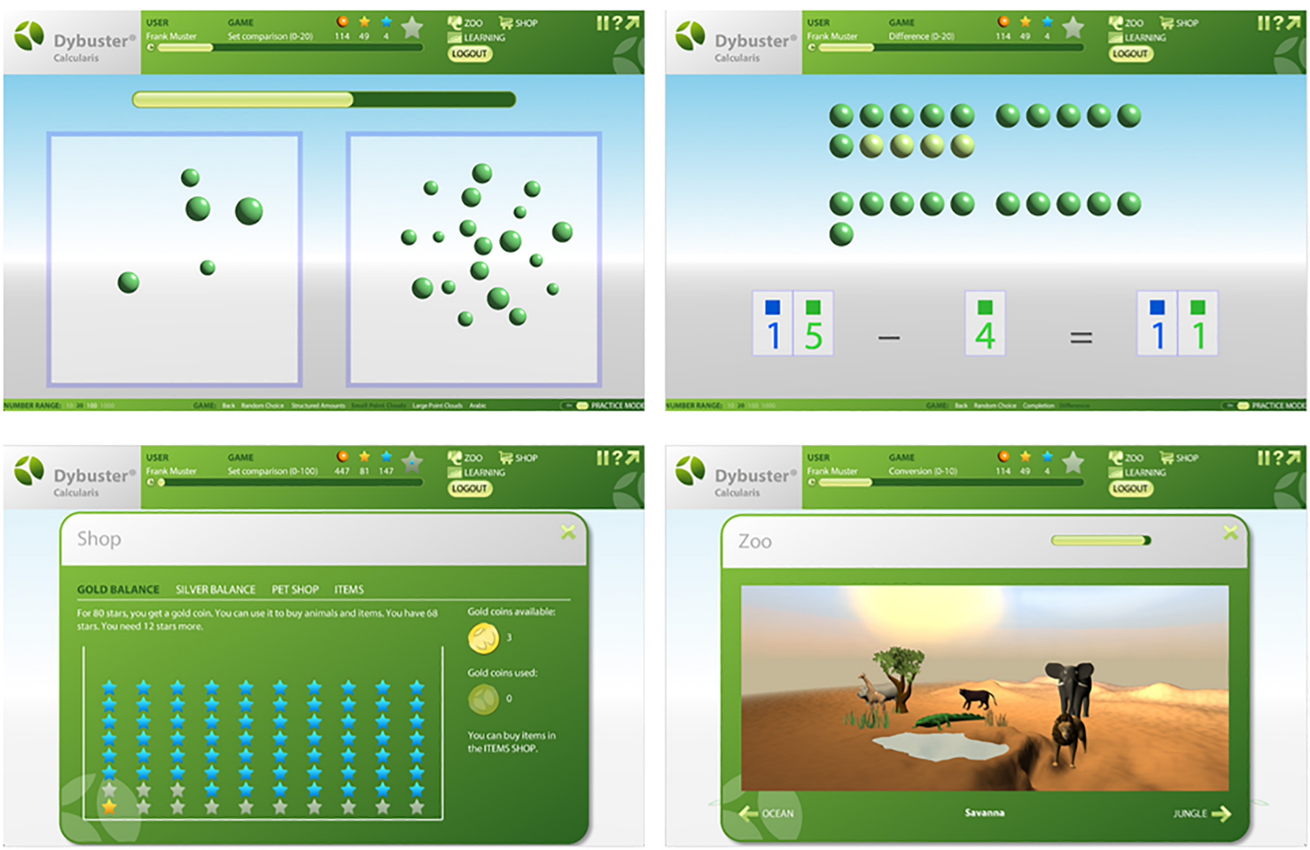


FIGURE 1 | Screenshots from the computer-based training program *Calcularis 2.0*. **(Top)** Left: magnitude comparison non-structured stimuli, right: subtraction with balls. **(Bottom)** Left: reward system – shop to buy and feed animals which can be assigned to various zoo worlds, right: zoo world Savanna.

power component and the IQ-score was within the normal range ($T \geq 40$) (Basic Diagnostics of specific developmental disorders in elementary school age children, BUEGA, Esser et al., 2008). Children were recruited consecutively via three outpatient clinics as well as via pediatricians in Germany. This approach addressed children with arithmetic problems with and without comorbid disorders. To make sure that enough children fulfill the determined criteria of DD, 107 children were screened.

Children were randomly assigned to the Calcularis group or the control group. Children of the Calcularis group completed a 12 weeks training, while the control group received no training. Children of the control group performed the training between time 2 (t_2) and time 3 (t_3). Children of both groups attended regular schools and visited regular math classes.

Children of the Calcularis group trained with the program 4–5 times per week with training sessions of 20 min after school. Children were assessed before and after the 12-weeks period (t_1/t_2) to evaluate the immediate training effects. To determine the stability of the training effects, children of the Calcularis group were re-assessed after a 3-months-interval (t_3).

Initial diagnostic included the assessment of mathematic competencies (*RZD*, Jacobs and Petermann, 2005) as well as intelligence (*BUEGA*, Esser et al., 2008) and math anxiety (*Math anxiety interview*, *MAI*, Kohn et al., 2013). The pre-/post-/follow-up test diagnostic ($t_1/t_{2/}t_3$) for children of both groups included the assessment of arithmetic performance (*Heidelberger Rechentest 1–4*, *HRT*, Haffner et al., 2005), reading and spelling (*BUEGA*, Esser et al., 2008), spatial representation of numbers (number line test 0–100) and basic number processing (basic number processing computer test).

Seventy-two German-speaking children could be included in the study (Calcularis group: $n = 39$, control group: $n = 33$). Only children with at least 42 sessions (corresponds to 70% of the maximum of 60 sessions) of Calcularis within a maximum of 13 weeks of training were included in the analysis. Due to these training-related inclusion criteria as well as other reasons such as illness during the training or test sessions, five children from the Calcularis group were excluded. The final study sample consisted of 67 children between the ages of 7.0–10.11 years attending second to the fifth grade of elementary school. The study population involved more girls ($n = 49$) than boys ($n = 18$), but gender ratio deviated not significantly over the groups.

Instruments

Basic Diagnostics of Specific Developmental Disorders in Elementary School Age Children (BUEGA)

The BUEGA (Esser et al., 2008) served for the assessment of verbal and non-verbal intelligence as well as the performance in reading, spelling, and arithmetic. The internal consistency coefficients determined for each school grade are sufficient to high ($\alpha = 0.81$ to $\alpha = 0.95$). The combined score for the reading and spelling performance is the mean value of the scores (standardized T-scores) achieved in reading and spelling.

Rechenfertigkeiten- und Zahlenverarbeitungs-Diagnostikum for the 2nd to 6th Grade (RZD 2–6)

The RZD 2–6 (Jacobs and Petermann, 2005) is a standardized mathematics test for diagnosing DD. The test assesses basic numerical capacities (e.g., transcoding, counting, number/quantity comparison, and spatial number representation) as well as arithmetic skills (addition, subtraction, multiplication, and division). The test allows for a differentiated assessment of the task performance of the child (power component) and the child's required time to solve the tasks (speed component). The reliability coefficients of both components (power component: $\alpha = 0.89$ to $\alpha = 0.90$; speed component: $\alpha = 0.89$ to $\alpha = 0.92$) are sufficient to high.

Math Anxiety Interview (MAI)

The MAI (Kohn et al., 2013) served to assess the children's math anxiety with the help of an anxiety thermometer. The children were asked to rate their intensity of math anxiety in four different situations which were illustrated with pictures. To rate their intensity, they got a thermometer made of cardboard, where they could adjust their fear by manually moving the red column in the thermometer from no anxiety at all or a lot of anxiety. Internal consistency measured using Cronbach's Alpha is sufficient ($\alpha = 0.76$).

Heidelberger Rechentest 1–4 (HRT)

The scale "arithmetic operations" of the HRT (Haffner et al., 2005) served to assess the children's arithmetic performance. The scale consists of six subtests (addition, subtraction, multiplication, and division as well as two further subtests with a slightly more complex format: Complete the task by filling in the missing number, e.g., 3 + ? = 5 or put the appropriate relation sign [>, <, =] in the box to show which number [left or right] is larger or if both are equal, e.g., 5 − 1 ? 4).

The HRT is designed as a speed test and specifically addresses computational fluency. For each subtest a score is determined based on the number of correctly solved items within the 2-min time limit This score is converted into a T-Score (based on norm values), subsequently the six T-Scores are added and in turn converted into a T-score for the entire scale.

As an index of reliability, retest reliability was calculated over a 2-week period with medium to high coefficients for the subtests ($r_{tt} = 0.77$ to $r_{tt} = 0.89$) as well as the over-all scale score ($r_{tt} = 0.93$).

Number Line Test

As a measure for spatial representation of numbers a number line test from 0 to 100 was administered. Children indicate the location of 20 verbally and visually presented numbers on a number line from 0 to 100. The percent absolute estimation error (PAE) for the target number and the indicated location (estimated number) on the number line was calculated (PAE = | estimated number–target number| /scale of estimates, cf. Siegler and Booth, 2004). In addition, to evaluate the linearity of the spatial representation we calculated the correlation coefficient of linear fit (R^2_{lin}) for each child (higher value is associated with better performance). Reliability coefficients estimated for PAE were sufficiently high (α = 0.81 to α = 0.94).

Basic Number Processing Computer Test

The subtests *single-digit number comparison*, *two-digit number comparison* and *magnitude comparison* of the computerized test battery of Landerl (2013) served as a measure of basic number processing. In the number comparison subtests (single-digit and two-digit numbers), children were presented with pairs of yellow digits on the computer screen and were asked to select the numerically larger one by pressing the corresponding keyboard button. In the single-digit task, 56 trials with numerical distances from 1 to 8 (36 trials for distance 1–3 and 20 trials for distance 4–8) were presented.

In the 2-digit task 80 trials were presented. To control for a unit-decade-compatibility effect (Nuerk et al., 2004), the influence of differences in the magnitude of decade and unit should be balanced. Therefore, 30 compatible (both decade and unit of one number are larger than decade and unit of the other, so decade and unit comparisons lead to the same response, e.g., 25 36), 30 incompatible (decade and unit comparisons led to different responses, e.g., 25 19), and 20 neutral items (both decades are the same, e.g., 25 29) were presented.

In the magnitude comparison task two quantities of randomly arranged yellow squares (20–72) were presented on the screen and children were supposed to select the numerically larger quantity. Out of the 57 trials there were 27 with a small distance (8–16) and 30 trials with a large distance (17–25).

Reaction times and errors were recorded by the computer. Reliability coefficients estimated for reaction times at each assessment point were high (single-digit: α = 0.95 to α = 0.96, two-digit: α = 0.95 to α = 0.96, magnitude comparison: α = 0.90 to α = 0.94).

The proportions of the correctly solved tasks (accuracy) as well as the individual median reaction times (for correct answers within a range of 200 ms to 10,000 ms) were calculated for each child. According to Landerl (2013) both measures (accuracy and speed of response) were combined into one measure, the inverse efficiency (IE), by dividing the median reaction times by the proportion of correct responses.

Statistical Analyses

Group differences were analyzed using Analyses of Variance (ANOVA) and Chi-square tests. A series of repeated measures general linear model (GLM) analyses as well as t-tests for paired samples were conducted to evaluate training effects between assessment time points ($t_1 - t_2$) as a within-subject factor and group (Calcularis group/control group) as a between-subject factor. The group x time interaction was the primary effect of interest. Effect sizes are expressed as partial eta squared (η^2) coefficients. Cohen (1988) postulates that η^2 values between 0.06 and 0.13 are medium effects and η^2 values greater than 0.14 are large effects. Correlation analyses and hierarchical regressions were applied to determine the effects of baseline factors on the individual response to the training.

RESULTS

The analyzed sample consisted of 67 children with developmental dyscalculia. The mean age was 8.96 (SD = 0.82) years. Children of the Calcularis group trained with the program for an average training duration of 11.47 (SD = 0.93) weeks and completed on average 53.29 (SD = 5.45, 42–62) training sessions. Statistical analyses revealed no significant differences between the groups for gender, age, arithmetic/numerical performance or control variables (intelligence, spelling, reading, additional reading, and/or spelling disorder) in the initial diagnostic procedure (t_1) (see **Table 1**). Criteria for a reading and/or spelling disorder were met if a child's performance in reading (composite of reading speed and accuracy BUEGA) or spelling (grapheme score BUEGA) was 1.5 standard deviations below the average ($T \leq 35$).

Immediate Training Effects

The mean values of the pre- and post-test scores regarding arithmetic performance, basic numerical processing and reading and spelling performance are presented in **Table 2**.

HRT

The repeated-measures GLM for the HRT "arithmetic operations" demonstrated a significant main effect of time (η^2 = 0.16), but no main effect of group. The group × time interaction was significant with medium effect size (η^2 = 0.10), indicating that training progress differed between both groups over time. Children of the Calcularis group demonstrated stronger improvements [$t(33) = -4.32$, $p < 0.001$] than the control group [$t(32) = -0.59$, $p = 0.559$].

Number Line Test

The results of the number line test with regard to PAE revealed a significant main effect of time (η^2 = 0.10). The group × time interaction was not significant. There was no main effect of group.

TABLE 1 | Demographic and cognitive characteristics [*Mean* (*SD*)] of the Calcularis group (CAL) and the control group (CG) prior to the intervention (t_1).

	CAL (*n* = 34)	CG (*n* = 33)	Test statistic	*p*
Gender (f/m)	26/8	23/10	0.39[d]	0.532
Age (years)	8.94 (0.77)	8.98 (0.88)	−0.22[e]	0.830
Calculation power component[a] (RZD)	34.20 (7.48)	33.70 (6.41)	0.29[e]	0.770
Calculation speed component[a] (RZD)	29.94[b] (3.91)	30.76[c] (4.35)	−0.71[e]	0.478
Mathematical performance[a] (BUEGA)	35.79 (7.90)	38.21 (7.61)	−1.28[e]	0.207
Intelligence[a] (BUEGA)	49.18 (6.90)	48.82 (6.38)	0.22[e]	0.826
Reading and spelling[a] (BUEGA)	40.04 (8.03)	41.00 (8.23)	−0.48[e]	0.632
Reading and/or spelling disorder	18 (52.9%)	13 (39.4%)	1.24[d]	0.266

[a] *T-score, RZD speed component not determinable in case of no correct item in one of the subtests leading to reduced sample sizes,* [b] *n = 25,* [c] *n = 27,* [d] *χ^2 Score,* [e] *t-score.*

TABLE 2 | Training effects (mean values and standard deviations) of the Calcularis group (CAL) and the control group (CG) in arithmetic performance, spatial number representation, basic numeric processing and reading and spelling.

Outcome parameter	Group	*n*	t_1	t_2	Effects	*F*	*p*	η^2
			M (SD)	*M (SD)*				
Arithmetic operations[a] (HRT)	CAL	34	31.35 (5.07)	34.68 (6.27)	Time	12.64	0.001	0.163
	CG	33	32.88 (6.75)	33.30 (6.78)	Group	0.003	0.958	0.000
					Group × Time	7.57	0.008	0.104
Number line test 0–100 (PAE)[b]	CAL	34	7.87 (3.31)	5.74 (2.56)	Time	7.12	0.010	0.099
	CG	33	8.69 (5.25)	8.30 (4.28)	Group	3.99	0.050	0.058
					Group × Time	3.38	0.070	0.049
Number line test 0–100 (R^2_{lin})	CAL	34	0.86 (0.11)	0.93 (0.07)	Time	7.01	0.010	0.097
	CG	33	0.85 (0.15)	0.85 (0.19)	Group	2.32	0.133	0.034
					Group × Time	5.52	0.022	0.078
1-digit comparison, IES (ms)[c]	CAL	31	954.21 (227.96)	819.22 (171.64)	Time	31.70	0.000	0.342
	CG	32	959.54 (166.39)	877.56 (202.77)	Group	0.50	0.480	0.008
					Group × Time	1.89	0.174	0.030
2-digit comparison, IES (ms)[c]	CAL	30	1868.80 (546.17)	1648.90 (461.76)	Time	5.45	0.023	0.083
	CG	32	1998.53 (599.87)	1891.97 (639.96)	Group	2.18	0.145	0.035
					Group × Time	0.66	0.421	0.011
Quantity comparison, IES (ms)[c]	CAL	32	1086.83 (265.19)	871.53 (194.31)	Time	65.63	0.000	0.514
	CG	32	1085.43 (211.27)	976.21 (189.11)	Group	1.05	0.310	0.017
					Group × Time	7.01	0.010	0.102
Reading and spelling[a] (BUEGA)	CAL	34	40.04 (8.02)	40.51 (7.98)	Time	0.33	0.566	0.005
	CG	33	41.00 (8.23)	39.86 (8.16)	Group	0.01	0.936	0.000
					Group × Time	1.94	0.168	0.029

[a]*T-score,* [b]*distance (percentage) from correct position,* [c]*inverse efficiency score.*

With regard to linearity, the group × time interaction was significant with moderate effect size (η^2 = 0.08), demonstrating stronger improvements for the Calcularis group [$t(33) = -4.33$, $p < 0.001$] compared to the CG [$t(32) = -0.18$, $p = 0.857$]. There was a significant main effect of time (η^2 = 0.10), but no main effect of group.

Basic Number Processing Computer Test

The analyses for the 1-digit comparison (IES) indicated a significant main effect of time (η^2 = 0.342), but no effect of group. The group × time interaction was not significant.

The analyses for the 2-digit comparison indicated a significant main effect of time (η^2 = 0.083), but no effect of group. The group × time interaction was not significant.

Regarding the IES of the quantity comparison task there was a significant group × time interaction (η^2 = 0.10). Children of the Calcularis group demonstrated stronger gains than the control group with medium effect size. The significant main effect of time (η^2 = 0.51) shows that both groups improved with regard to IES but the Calcularis group [$t(31) = 8.14$, $p < 0.001$] outperformed the control group [$t(31) = 3.63$, $p = 0.001$]. No significant main effect of group was found.

Reading and Spelling Performance

As a measure of domain specificity, the reading and spelling performance was assessed, and the mean of both measures was used as the dependent variable. The analysis yielded no main effects of time, nor group. The interaction between group x time was not significant for the comparison between the Calcularis and the control group.

To summarize, group × time effects were found for the arithmetic operations (HRT), linearity of the number line and quantity comparison tasks, but not for the score PAE (number line task) and the number comparison tasks, implying that the Calcularis group improved on arithmetic performance (including addition and subtraction), spatial number processing and magnitude comparison.

Stability of the Training Effects

The analysis of the stability of the training effects ($t_2 - t_3$) refers only to the Calcularis group since the control group served as a waiting control group and received the computerized training during this interval ($t_2 - t_3$). The results concerning the stability of the training effects demonstrate that the Calcularis group showed moderate to high correlation coefficients (r = 0.59 to r = 0.88) for all measures of basic numerical processing and arithmetic competencies. The paired samples t-tests revealed no significant results demonstrating stable training effects after a 3-months-interval (see **Table 3**), with the exception of the number line test 0–100 (R^2_{lin}). Children showed reduced scores in linearity (R^2_{lin}) while the scores were still significantly higher than at the beginning of the training [t_1: $M = 0.86$, $SD = 0.11$, t_3: $M = 0.90$, $SD = 0.12$, $t(31) = -2.14$, $p = 0.041$].

Factors Predicting Training Gain

To investigate whether baseline measures predict individual differences in training improvement, we examined the relation between postulated baseline measures and changes in arithmetic performance as the most curriculum-related criterion (HRT arithmetic operations t_2 minus HRT arithmetic operations t_1).

TABLE 3 | Stability of training effects of the Calcularis group in arithmetic performance, spatial number representation, basic numeric processing and reading and spelling (mean values and standard deviations for t_2 and t_3), correlation coefficients *r* and *t*-tests.

Outcome parameter	*n*	t_2	t_3	Correlation *r*	*t*-test	
		M (SD)	*M (SD)*		*t*	*p*
Arithmetic operations[a] (HRT)	32	34.91 (6.39)	35.19 (6.37)	0.77	−0.37	0.716
Number line test 0–100 (PAE)[b]	32	5.76 (2.64)	6.19 (2.70)	0.66	−1.10	0.281
Number line test 0–100 (R^2_{lin})	32	0.93 (0.07)	0.90 (0.12)	0.64	2.43	0.021
1-digit comparison, IES (ms)[c]	30	815.18 (176.44)	791.01 (192.08)	0.68	0.90	0.377
2-digit comparison, IES (ms)[c]	28	1634.47 (474.35)	1558.51 (430.32)	0.78	1.31	0.200
Quantity comparison, IES (ms)[c]	31	868.76 (197.10)	886.93 (229.07)	0.59	−0.52	0.606
Reading and spelling[a] (BUEGA)	32	40.67 (8.08)	41.05 (9.31)	0.88	−0.48	0.635

[a]*T-score,* [b]*distance (percentage) from correct position,* [c]*inverse efficiency score.*

Results are presented in **Table 4**, showing significant negative correlation coefficients between arithmetic improvement and math anxiety ($r = -0.35$, $p = 0.020$) and an additional reading/spelling disorder ($r = -0.43$, $p = 0.005$). Additionally, there was a small correlation coefficient between arithmetic improvement and general intelligence (t_1) $r = 0.25$, $p = 0.074$, but no significant correlations between arithmetic improvement and number of sessions or Arithmetic operations (t_1).

To examine which of the baseline measures predicted unique variance in mathematics achievement scores (gain) a hierarchical regression analysis was conducted. Independent variables were added in a stepwise procedure. This method allowed to control for general intelligence (t_1) (step 1), before investigating the unique contribution of the potential predictors in step 2 (additional reading/spelling disorder, t_1) and step 3 (math anxiety, t_1) to the variance in arithmetic improvement. Results from this model (see **Table 5**) demonstrated that an additional reading/spelling disorder explained a significant amount of unique variance in arithmetic improvement [$\Delta R^2 = 0.17$, $F(1,31) = 7.05$, $p = 0.012$]. The negative standardized beta-coefficient as well as the negative correlation coefficient indicated that children with an additional reading/spelling disorder show smaller improvements. Additionally, math anxiety also explains a significant amount of unique variance in arithmetic improvement [$\Delta R^2 = 0.12$, $F(1,30) = 5.48$, $p = 0.026$]. The negative beta weight indicated that children with higher math anxiety show less improvement.

TABLE 4 | Correlations among predictor measures (t_1) and gain ($t_2 - t_1$) ($n = 34$).

	1	2	3	4	5	6
(1) Gain (arithmetic operations, HRT)	–	−0.074	0.254⁺	−0.354*	−0.433**	−0.097
(2) Arithmetic operations (HRT)[a] (t_1)		–	0.064	−0.236⁺	−0.151	−0.222
(3) Intelligence (BUEGA)[a] (t_1)			–	0.002	−0.071	0.354*
(4) Math anxiety (MAI) (t_1)				–	0.027	0.142
(5) Reading/spelling disorder (t_1)					–	0.161
(6) Number of sessions						–

⁺*p < 0.010, *p < 0.05, **p < 0.01, n = 34,* [a]*T-score.*

DISCUSSION

The aim of the present study was the evaluation of the adaptive computer-based training program *Calcularis 2.0* in a sample of dyscalculic children. Furthermore, factors that predict training improvement were investigated.

Immediate Training Effects

As expected, compared to the (waiting) control group, the Calcularis group demonstrated larger improvements with moderate effect sizes in a standardized math achievement test (HRT) ($g = 0.49$), in spatial number processing ($g = 0.55$) and magnitude comparison ($g = 0.44$). No training effects were found for reading and spelling performance, hence the presented findings can be interpreted as an indicator for domain specificity of the training.

The HRT is designed as a speed test and specifically addresses arithmetic fluency. It is assumed that the training leads to a higher automation of task processing resulting in faster fact retrieval. Compared to the evaluative studies regarding the previous version *Calcularis 1.0* (Käser et al., 2013a; Rauscher et al., 2016; Kohn et al., 2017) the observed training effects are stronger, whereby it has to be considered that *Calcularis 2.0* includes additional tasks and additional motivational components and that the training interval was prolonged. The medium effect sizes are comparable to other trainings (Kroesbergen and van Luit, 2003; Ise et al., 2012; Chodura et al., 2015) and are satisfactory for a sample of children with severe deficits (participants with DD).

Regarding spatial number processing (number line test) the Calcularis group showed a significant decrease in PAE and a significant increase in linearity, but only the change in linearity was significantly higher than in the control group. These findings are in line with a previous study (Käser et al., 2013a) that analyzed PAE and showed an improvement for the number range 0–100 after a 3-months training period. These results are promising, as the mastering of number line tasks constitutes an important step in the numerical development (von Aster and Shalev, 2007) and provides a tool for solving basic arithmetic. However, it must

TABLE 5 | Hierarchical regression analysis for the prediction of gain (arithmetic operations, HRT, $t_2 - t_1$, $n = 34$).

Variable	R^2	ΔR^2	ΔF	Standardized β	t	p
Step 1	0.064		(1,32) = 2.21, p = 0.147			
Intelligence t_1				0.254	1.485	0.147
Step 2	0.238	0.173	(1,31) = 7.05, p = 0.012			
Intelligence t_1				0.224	1.427	0.163
Reading/spelling disorder t_1				−0.417	−2.656	0.012
Step 3	0.355	0.118	(1,30) = 5.48, p = 0.026			
Intelligence t_1				0.226	1.537	0.135
Reading/spelling disorder t_1				−0.408	−2.776	0.009
Math anxiety t_1				−0.343	−2.340	0.026

be taken into account that this improvement on the number line task might not only be due to an improvement of this underlying mental number line. Recent studies indicate that the improvement could also reflect an increasing use of helpful strategies, like using reference points at the number line (Ashcraft and Moore, 2012; Link et al., 2014; Peeters et al., 2016).

With respect to basic number processing no training effects were found for number comparison (1-digit/2-digit), but for magnitude comparison. Compared to the control group, the Calcularis group demonstrated larger improvements with moderate effect size. The low baseline level is one possible explanation of these non-expected results regarding number comparison. Compared to the findings of Landerl (2013) who used the same experimental design, the observed inverse efficiency scores (ms) in our study were lower (i.e., better), providing less room for improvement. Furthermore, both groups (CAL and CG) demonstrated decreased IE-scores that may indicate a test repetition effect, so the additional improvement through the training could possibly not be observed (see **Table 2** for 1-digit and 2-digit comparison).

Furthermore, the concept of the training program has to be taken in consideration which balances the training time between the area of number representations and arithmetic operations. Additionally, there is a high variety of skills that are trained in the area of number representations. Therefore, some skills are only trained for a short amount of time or especially at the beginning of the training. As mentioned above children are considered to be already quite proficient in 1-digit comparison and because of the highly adaptive training program, the training sequence of this skill was passed rapidly. However, the result concerning magnitude comparison is promising since faster reaction times in symbolic as well as non-symbolic comparisons are related to higher calculation fluency (for detailed review see De Smedt et al., 2013). Of course, it has to be pointed out that this relationship should rather be interpreted bidirectionally and not causally.

Stability of the Training Effects

Children were re-assessed after a 3-months interval to determine the stability of the training effects. Regarding all measures of basic numerical processing and arithmetic competencies results demonstrated stable performance scores with moderate to high correlation coefficients that indicates that the children keep their relative position. The performance improvements of the intervention ($t_1 - t_2$) were shown to be stable after a 3-months-interval ($t_2 - t_3$). Only the linearity results (R^2_{lin}) showed a significant decrease ($t_2 - t_3$), although the scores were still significantly higher than at the beginning of the training. It has to be mentioned that a comparison to a group without any intervention from t_1 to t_3 is missing because we were unable to include a waiting control group over 6 months due to ethical reasons. Therefore, we were not able to control for developmental effects and the results provide merely indirect evidence for stable training effects.

The found follow-up effects are comparable to Fischer et al. (2008) and even better than in a former study evaluating Calcularis 1.0 (Kohn et al., 2017). Furthermore, the results support the assumption that a prolonged training duration (12 weeks in Calcularis 2.0 instead of 6 weeks in Calcularis 1.0) contributes to more robust effects.

Factors Predicting Training Improvement

A hierarchical regression analysis indicates that dyscalculic children without an additional reading/spelling disorder as well as those with low math anxiety scores show higher improvement scores.

It is assumed that children suffering from math anxiety tend to avoid math-related tasks (Ashcraft et al., 2007). We therefore suppose that throughout the training high anxious children tended to confront themselves less with the offered tasks, tended to demonstrate less elaborated processing of the content, and tended to show more off-task behavior.

Therefore, they did not improve their achievement scores as much as their non-anxious peers. This assumption could be integrated in the debilitating anxiety model (Carey et al., 2016). As there were no significant differences between low anxious (n = 16, M = 31.87, SD = 4.33) and high anxious children (n = 17, M = 30.29, SD = 3.80) with respect to arithmetic performance at baseline [$t(31) = 1.12$, $p = 0.273$], the hypothesis that math anxiety inhibits a significant improvement should be verified in further analyses. A first look at the number of training sessions indicated that there is no easy answer to this question. There was no significant difference [$t(31) = -1.16$, $p = 0.257$] between high anxious (n = 16,

M = 54.13, SD = 5.39) and low anxious children (n = 17, M = 52.00, SD = 5.17) concerning the number of training sessions. Accordingly, the assumption of a different training behavior must be analyzed based on the log data of each child. Due to the focus of this paper these questions should be elaborated in detail in a subsequent study. Before doing so, a theoretical and methodological clarification of the construct "off-task behavior" is necessary, which affects various aspects of the training behavior.

In line with previous research by Powell et al. (2009) we found a better responsiveness to the training for children without an additional reading/spelling disorder. It is assumed that children with a comorbid dyslexia show underlying phonological processing deficits (Hecht et al., 2001; Robinson et al., 2002) and greater deficits in verbal and visuospatial working memory (Swanson et al., 2009). Therefore, children with comorbid dyslexia could have additional problems that could not be addressed successfully in the 12-weeks-training.

Concerning the predictor of intellectual ability at baseline, the result was less substantial. There seemed to be a trend that DD-children with higher intelligence scores showed higher improvement scores. That would be in line with the results presented by Nemmi et al. (2016). In contrast to this study, initial arithmetic performance (t_1) did not predict individual responsiveness. That could be attributed to the fact that only children with DD were considered in the present study.

Limitations and Further Research Indications

When interpreting the findings, some methodological limitations must be considered.

First, the present study design includes the comparison to an untrained control group whereas comparisons to groups receiving alternative trainings are missing. The implementation of an untrained waiting control group allows the delineation of specific training effects to developmental and schooling effects. However, it has to be questioned whether these severely affected children in the untreated group actually did not receive additional support during the waiting period. Factors such as increased parental assistance in math exercises or a stronger response by teachers, as well as expectation effects (in the sense of a placebo effect) are conceivable. An alternative systematic treatment would have strengthened the findings and is therefore planned in future intervention studies.

Second, a high external validity of the clinical sample was required, leading to a high amount of comorbidities such as dyslexia and probably attention deficit hyperactivity disorders (Auerbach et al., 2008; Fischbach et al., 2010).

As this study and Rauscher et al. (2017) show that comorbid dyslexia can influence the results it is absolutely necessary to use a design with a larger sample that enables to compare children with single and combined deficits to replicate the promising effects. Including a larger sample in future training studies would also offer a higher statistical power and allow for a deeper analysis of the differential efficacy as well as for essential replications.

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. First, training a variety of skills shortens the training time of each specific skill and thus leads to smaller training effects as mentioned above.

Second, due to the high adaptability of the program, each child pursues a different training trajectory. Since it is not obvious, which aspects of the training lead to which performance improvement, modular tests or deeper analyses of the individual pathways could be beneficial.

CONCLUSION

This study demonstrates that the adaptive training program *Calcularis 2.0* can be used effectively to support dyscalculic children in their numerical achievement. The results showed that even after a rather short training period of 12 weeks, solid training effects with regard to arithmetic and spatial number representation could be achieved. Results indicate that especially math anxiety and a co-occurring reading and/or spelling disorder were significant predictors for individual responsiveness to this training. The training effects were shown to be stable after a 3-months-interval.

In practice, *Calcularis 2.0* can be used individually as well as in a group or class setting as a beneficial enhancement of learning intervention and math lessons. Based on the results of this evaluation and former results (Käser et al., 2013a), a training period of at least 3 months with a training frequency of 3–4 training sessions per week is recommended. The children can work on their own, without performance pressure and frightening peer comparisons. It is important to highlight that *Calcularis 2.0* aims not to be a substitute for teachers, since a positive learning development is created by educational skills, methodical knowledge and an encouraging teacher–student relationship. Especially for children with high math anxiety it should be considered that a one-to-one tutoring might be more effective and might address the individual experiences of fear in the former learning history to help to recode and overcome these internal representations. However, it also seems possible to develop and include elements for the detection of emotional states and for an according cognitive behavioral intervention into future learning environments (O'Neill and Gillespie, 2014).

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The study involving human participants was reviewed and approved by the Ethics Committee of the University of Potsdam.

Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

JK: conceptualization of the study idea and design, organization of the database, statistical analysis and interpretation of the data, and preparation and revision of the manuscript. LR: conceptualization of the study idea and design, organization of the database, collection and interpretation of the data, and preparation and revision of the manuscript. TK: conceptualization of study idea, writing sections of the manuscript, and revision of the manuscript. KK, AW, and GE: conceptualization of the study design and revision of the manuscript. MA: conceptualization of the study idea and design, interpretation of the data, writing sections of the manuscript, and revision of the manuscript. All authors have contributed and approved the final manuscript.

FUNDING

The present study was financially supported by a research grant of the German Federal Ministry of Education and Research (BMBF, numbers: 01GJ1011 and 01GJ1301).

ACKNOWLEDGMENTS

We acknowledge the support of the Deutsche Forschungsgemeinschaft and Open Access Publishing Fund of University of Potsdam.

REFERENCES

American Psychiatric Association (2013). *Diagnostic and Statistical Manual Of Mental Disorders*, 5th Edn, Washington, DC: American Psychiatric Publishing.

Ashcraft, M. H., and Faust, M. W. (1994). Mathematics anxiety and mental arithmetic performance: an exploratory investigation. *Cogn. Emot.* 8, 97–125. doi: 10.1080/02699939408408931

Ashcraft, M. H., and Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *J. Exper. Psychol. Gen.* 130, 224–237. doi: 10.1037/0096-3445.130.2.224

Ashcraft, M. H., Krause, J. A., and Hopko, D. R. (2007). "Is math anxiety a mathematical learning disability?," 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. Michèle (Baltimore, MD: Paul H. Brookes Pub. Co), 329–348.

Ashcraft, M. H., and Moore, A. M. (2009). Mathematics anxiety and the affective drop in performance. *J. Psychoeduc. Assess.* 27, 197–205. doi: 10.1177/0734282908330580

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

Auerbach, J. G., Gross-Tsur, V., Manor, O., and Shalev, R. S. (2008). Emotional and behavioral characteristics over a six-year period in youth with persistent and nonpersistent dyscalculia. *J. Learn. Disabil.* 41, 263–273. doi: 10.1177/0022219408315637

Baker, S., Gersten, R., and Lee, D. (2002). A synthesis of empirical research on teaching mathematics to low-achieving students. *Element. Sch. J.* 103, 51–73. doi: 10.1086/499715

Bryant, D. P., Bryant, B. R., Gersten, R., Scammacca, N., and Chavez, M. M. (2008). Mathematics instruction for first- and second-grade students with mathematics difficulties. *Remed. Spec. Educ.* 29, 20–32. doi: 10.1177/0741932507309712

Burns, M. K., Codding, R. S., Boice, C. H., and Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instructional and frustration level skills: evidence for a skill-by-treatment interaction. *Sch. Psychol. Rev.* 39, 69–83.

Butterworth, B., and Laurillard, D. (2010). Low numeracy and dyscalculia: identification and intervention. *ZDM Math. Educ.* 42, 527–539. doi: 10.1007/s11858-010-0267-4

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

Carey, E., Hill, F., Devine, A., and Szücs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Front. Psychol.* 6:1987. doi: 10.3389/fpsyg.2015.01987

Chodura, S., Kuhn, J. T., and Holling, H. (2015). Interventions for children with mathematical difficulties: A meta-analysis. *Zeitschr. Psychol.* 223, 129–144. doi: 10.1027/2151-2604/a000211

Cohen, J. (1988). *Statistical Power Analysis For The Behavioral Sciences*. New Jersey: Lawrence Erlbaum Associates.

Cohen Kadosh, R., Dowker, A., Heine, A., Kaufmann, L., and Kucian, K. (2013). Interventions for improving numerical abilities: Present and future. *Trends Neurosci. Educ.* 2, 85–93. doi: 10.1016/j.tine.2013.04.001

De Smedt, B., Noël, M.-P., Gilmore, C., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills 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

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-n

Di Giorgio, E., Lunghi, M., Rugani, R., Regolin, L., and Dalla Barba, B. (2019). A mental number line in human newborns. *Dev. Sci.* 22:e12801. doi: 10.1111/desc.12801

Dowker, A. (2004). *What Works For Children With Mathematical Difficulties? Research Report RR554*. Oxford: University of Oxford.

Dowker, A. (2007). What can intervention tell us about arithmetical difficulties? *Educ. Child Psychol.* 24, 64–75.

Dowker, A., Bennett, K., and Smith, L. (2012). Attitudes to mathematics in primary school children. *Child Dev. Res.* 2012:124939. doi: 10.1155/2012/124939

Esser, G., Wyschkon, A., and Ballaschk, K. (2008). *Basisdiagnostik Umschriebener Entwicklungsstörungen im Grundschulalter (BUEGA)*. Göttingen: Hogrefe.

Fischbach, A., Schuchardt, K., Brandenburg, J., Klesczewski, J., Balke-Melcher, C., Schmidt, C., et al. (2013). Prävalenz von Lernschwächen und Lernstörungen: Zur Bedeutung der Diagnosekriterien. *Lernen. Lernstörung.* 2, 65–76.

Fischbach, A., Schuchardt, K., Mähler, C., and Hasselhorn, M. (2010). Zeigen Kinder mit schulischen Minderleistungen sozio-emotionale Auffälligkeiten? *Zeitschr. Entwicklungspsychol. Pädagog. Psychol.* 42, 201–210. doi: 10.1026/0049-8637/a000025

Fischer, B., Köngeter, A., and Hartnegg, K. (2008). Effects of daily practice on subitizing, visual counting, and basic arithmetic skills. *Optom. Vis. Dev.* 39, 30–34.

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

Fuchs, L. S., Fuchs, D., Powell, S. R., Seethaler, P. M., Cirino, P. T., and Fletcher, J. M. (2008). Intensive intervention for students with mathematics disabilities:

Seven principles of effective practice. *Learn. Disabil. Q.* 31, 79–92. doi: 10.2307/20528819

Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., et al. (2010). The effects of strategic counting instruction, with and without deliberate practice, on number combination skill among students with mathematics difficulties. *Learn. Individ. Differ.* 20, 89–100. doi: 10.1016/j.lindif.2009.09.003

Geary, D. C., Hamson, C. O., and Hoard, M. K. (2000). Numerical and arithmetical cognition: a longitudinal study of process and concept deficits in children with learning disability. *J. Exp. Child Psychol.* 77, 236–263. doi: 10.1006/jecp.2000.2561

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

Haffner, J., Baro, K., Parzer, P., and Resch, F. (2005). *Heidelberger Rechentest (HRT): Erfassung mathematischer Basiskompetenzen im Grundschulalter*. Göttingen: Hogrefe.

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

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *J. Res. Math. Educ.* 21, 33–46.

Hillmayr, D., Reinhold, F., Ziernwald, L., and Reiss, K. (2017). *Digitale Medien im Mathematisch-Naturwissenschaftlichen Unterricht der Sekundarstufe: Einsatzmöglichkeiten, Umsetzung und Wirksamkeit*. Münster: Waxmann Verlag.

Ise, E., Dolle, K., Pixner, S., and Schulte-Körne, G. (2012). Effektive Förderung rechenschwacher Kinder: Eine Metaanalyse. *Kindheit Entwicklung* 21, 181–192. doi: 10.1026/0942-5403/a000083

Jacobs, C., and Petermann, F. (2005). *Rechenfertigkeiten- und Zahlenverarbeitungs-Diagnostikum für die 2. bis 6. Klasse*. Göttingen: Hogrefe.

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: The MIT Press.

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

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

Käser, T., Busetto, A. G., Solenthaler, B., Baschera, G. M., Kohn, J., Kucian, K., et al. (2013b). Modelling and optimizing mathematics learning in children. *Intern. J. Artif. Intellig. Educ.* 23:115.

Käser, T., Busetto, A. G., Baschera, G.-M., Kohn, J., Kucian, K., von Aster, M., et al. (2012). "Modelling and optimizing the process of learning mathematics," in *Proceedings of the 11th International Conference on Intelligent Tutoring Systems*, eds S. A. Cerri, W. J. Clancey, G. Papadourakis, and K. Panourgia (Heidelberg: Springer), 389–398. doi: 10.1007/978-3-642-30950-2_50

Käser, T., and von Aster, M. G. (2013). "Computerbasierte Lernprogramme für Kinder mit Rechenschwäche," in *Rechenstörungen bei Kindern: Neurowissenschaft, Psychologie, Pädagogik*, eds M. G. von Aster, and J. H. Lorenz (Göttingen: Vandenhoeck & Ruprecht), 259–275.

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.* 36, 564–573. doi: 10.1177/00222194030360060701

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

Kaufmann, L., and von Aster, M. (2012). The diagnosis and management of dyscalculia. *Deutsches Ärzteblatt Intern.* 109, 767–778.

Kaufmann, L., Wood, G., Rubinsten, O., and Henik, A. (2011). Meta-analyses of developmental 821 fMRI studies investigating typical and atypical trajectories of number processing and calculation. *Dev. Neuropsychol.* 36, 763–787. doi: 10.1080/87565641.2010.549884

Kohn, J., Rauscher, L., Käser, T., Kucian, K., McCaskey, U., Wyschkon, A., et al. (2017). Effekte des Calcularis-Trainings - Teil 1: Domänen-spezifische Veränderungen. *Lernen. Lernstörung.* 6, 51–63. doi: 10.1024/2235-0977/a000166

Kohn, J., Richtmann, V., Rauscher, L., Kucian, K., Käser, T., Grond, U., et al. (2013). Das Mathematikangstinterview (MAI): Erste psychometrische Gütekriterien. *Lernen. Lernstörung.* 2, 1–14.

Krajewski, K., and Ennemoser, M. (2010). "Die Berücksichtigung begrenzter Arbeitsgedächtnisressourcen in Unterricht und Lernförderung," in *Brennpunkte der Gedächtnisforschung*, eds H.-P. Trolldenier, W. Lenhard, and P. Marx (Göttingen: Hogrefe), 337–365.

Krinzinger, H., and Kaufmann, L. (2006). Rechenangst und Rechenleistung. *Sprache Stimme Gehör.* 30, 160–164. doi: 10.1055/s-2006-951753

Kroesbergen, E. H., and van Luit, J. E. H. (2003). Mathematics intervention for children with special educational needs: a meta-analysis. *Remed. Spec. Educ.* 24, 97–114. doi: 10.1177/07419325030240020501

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.* 32:340. doi: 10.1017/s0140525x09990069

Kucian, K., McCaskey, U., O'Gorman Tuura, R., and von Aster, M. (2018a). Neurostructural correlate of math anxiety in the brain of children. *Nat. Transl. Psychiatry* 8:273. doi: 10.1038/s41398-018-0320-326

Kucian, K., McCaskey, U., von Aster, M. G., and O'Gorman Tuura, R. (2018b). Development of a possible general magnitude system for number and space. *Front. Psychol.* 9:2221. doi: 10.3389/fpsyg.2018.02221

Kuhn, J.-T., and Holling, H. (2014). Number sense or working memory? The effect of two computer-based trainings on mathematical skills in elementary school. *Adv. Cogn. Psychol.* 10, 59–67. doi: 10.5709/acp-0157-2

Kuhn, J. T., Raddatz, J., Holling, H., and Dobel, C. (2013). Dyskalkulie vs. Rechenschwäche: Basisnumerische Verarbeitung in der Grundschule. *Lernen. Lernstörung.* 2, 229–247. doi: 10.1024/2235-0977/a000044

Kulik, C.-L. C., and Kulik, J. A. (1991). Effectiveness of computer-based instruction: an updated 837 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). "Computerunterstützte Rechentrainingsprogramme," in *Interventionen bei Lernstörungen*, eds W. Lauth, M. Grünke, and J. C. Brunstein (Göttingen: Hogrefe), 329–337.

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.2018.459

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

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

Link, T., Huber, S., Nuerk, H.-H., 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

Ma, X. (1999). A meta-analysis of the relationship between anxiety towards mathematics and achievement in mathematics. *J. Res. Math. Educ.* 30, 520–540.

Meece, J. L., Wigfield, A., and Eccles, J. S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrollment intentions and performance in mathematics. *J. Educ. Psychol.* 82, 60–70. doi: 10.1037/0022-0663.82.1.60

Moeller, K., Fischer, U., Cress, U., and Nuerk, H.-C. (2012). "Diagnostics and intervention in developmental dyscalculia: current issues and novel perspectives," in *Reading, Writing, Mathematics And The Developing Brain: Listening to Many Voices*, eds Z. Breznitz, O. Rubinsten, V. J. Molfese, and D. L. Molfese (New York, NY: Springer), 233–294.

Nemmi, F., Helander, E., Helenius, O., Almeida, R., Hassler, M., Räsänen, P., et al. (2016). Behavior and neuroimaging at baseline predict individual response to combined mathematical and working memory training in children. *Dev. Cogn. Neurosci.* 20, 43–51. doi: 10.1016/j.dcn.2016.06.004

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

O'Neill, B., and Gillespie, A. (eds) (2014). *Assistive Technology for Cognition(:)A Handbook For Clinicians And Developers*. Hove: Psychology Press.

Peeters, D., Degrande, T., Ebersbach, M., Verschaffel, L., and Luwel, K. (2016). Children's use of number line estimation strategies. *Eur. J. Psychol. Educ.* 31, 117–134. doi: 10.1007/s10212-015-0251-z

Powell, S. R., Fuchs, L. S., Fuchs, D., Cirino, P. T., and Fletcher, J. M. (2009). Effects of fact retrieval tutoring on third-grade students with math difficulties with and without reading difficulties. *Learn. Disabil. Res. Pract.* 24, 1–11. doi: 10.1111/j.1540-5826.2008.01272.x

Ramirez, G., Gunderson, E. A., Levine, S. C., and Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *J. Cogn. Dev.* 14, 187–202. doi: 10.1080/15248372.2012.664593

Räsänen, P. (2015). "Computer-assisted interventions on basic number skills," in *The Oxford Handbook Of Numerical Cognition*, eds A. Dowker, and R. Cohen Kadosh (Oxford: Oxford University Press), 745–766.

Räsänen, P., Käser, T., Wilson, A., von Aster, M., Maslov, O., and Maslova, U. (2015). "Assistive technology for supporting learning numeracy," in *Assistive Technology for Cognition: A Handbook for Clinicians and Developers*, eds B. O'Neill, and A. Gillespie (Glasgow: Psychology Press), 112–127.

Räsänen, P., Laurillard, D., Käser, T., and von Aster, M. (2019). "Perspectives to technology-enhanced-learning and teaching in mathematical learning difficulties," in *International Handbook Of Mathematical Learning Difficulties*, eds A. Fritz, V. G. Haase, and P. Räsänen (Berlin: Springer), 733–754. doi: 10.1007/978-3-319-97148-3_42

Räsänen, P., Saminen, J., Wilson, A. J., Aunio, P., and Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cogn. Dev.* 24, 450–472. doi: 10.1016/j.cogdev.2009.09.003

Rauscher, L., Kohn, J., Käser, T., Kucian, K., McCaskey, U., Wyschkon, A., et al. (2017). Effekte des Calcularis-Trainings - Teil 2: Veränderungen psychosozialer Merkmale. *Lernen. Lernstörung.* 6, 75–86. doi: 10.1024/2235-0977/a000168

Rauscher, L., Kohn, J., Käser, T., Mayer, V., Kucian, K., McCaskey, U., et al. (2016). Evaluation of a computer-based training program for enhancing arithmetic skills and spatial number representation in primary school children. *Front. Psychol.* 7:913. doi: 10.3389/fpsyg.2016.00913

Richardson, F. C., and Suinn, R. M. (1972). The mathematics anxiety rating scale: psychometric data. *J. Counsel. Psychol.* 19, 551–554. doi: 10.1037/h0033456

Robinson, C. S., Menchetti, B. M., and Torgesen, J. K. (2002). Towards a two-factor theory of one type of mathematics disabilities. *Learn. Disabil. Res.* 17, 81–89. doi: 10.1111/1540-5826.00035

Schoppek, W., and Tullis, 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

Seo, Y.-J., and Bryant, D. P. (2009). Analysis of studies of the effects of computer-assisted instruction on the mathematics performance of students with learning disabilities. *Comput. Educ.* 53, 913–928. doi: 10.1016/j.compedu.2009.05.002

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

Slavin, R., and Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Rev. Educ. Res.* 78, 427–515. doi: 10.3102/0034654308317473

Supekar, K., Iuculano, T., Chen, L., and Menon, V. (2015). Remediation of childhood math anxiety and associated neural circuits through cognitive tutoring. *J. Neurosci.* 35, 12574–12583. doi: 10.1523/jneurosci.0786-15.2015

Swanson, H. L., Jerman, O., and Zheng, X. (2009). Math disabilities and reading disabilities: can they be separated? *J. Psychoeduc. Assessm.* 27, 175–196. doi: 10.1177/0734282908330578

von Aster, M., Käser, T., Kohn, J., Kucian, K., Rauscher, L., and Vögeli, C. (2016). "Calcularis - Eine adaptive Lernsoftware zur Matheförderung," in *Förderprogramme für Vor- und Grundschule*, eds W. Schneider, and M. Hasselhorn (Göttingen: Hogrefe), 225–247.

von Aster, M., and Lipka, M. (2018). Droht der Schule ein digitaler Tsunami? - Auf die Qualität kommt es an! Lernen und Lernstörungen. *Science* 7, 65–66. doi: 10.1024/2235-0977/a000209

von Aster, M. G. (1994). Developmental dyscalculia in children: review of the literature and clinical validation. *Acta Paedopsychiatr.* 56, 169–178.

von Aster, M. G. (2000). Developmental cognitive neuropsychology of number processing and calculation: varieties of developmental dyscalculia. *Eur. Child Adolesc. Psychiatry* 9, 41–58.

von Aster, M. G., Käser, T., Kucian, K., and Gross, M. (2012). Calcularis - Rechenschwäche mit dem Computer begegnen. *Schweizer. Zeitschr. Heilpädag.* 18, 32–36.

von Aster, M. G., Rauscher, L., Kohn, J., and Eitel, Y. (2017). "Mathematikangst," in *Handbuch Rechenschwäche*, eds A. Fritz, G. Ricken, and S. Schmidt (Weinheim: Beltz).

von Aster, M. G., 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

Vukovic, R. K., Kieffer, M. J., Bailey, S. P., and Harari, R. R. (2013). Mathematics anxiety in young children: concurrent and longitudinal associations with mathematical performance. *Contemp. Educ. Psychol.* 38, 1–10. doi: 10.1016/j.cedpsych.2012.09.001

Wilson, A. J., Dehaene, S., Pinel, P., Revkin, S. K., Cohen, L., and Cohen, D. (2006). 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

Wu, S. S., Barth, M., Amin, H., Malcarne, V., and Menon, V. (2012). Math anxiety in second and third graders and its relation to mathematics achievement. *Front. Psychol.* 3:162. doi: 10.3389/fpsyg.2018.0162

Conflict of Interest: 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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 16 July 2020
doi: 10.3389/fnhum.2020.00271

Dissociating Arithmetic Operations in the Parietal Cortex Using 1 Hz Repetitive Transcranial Magnetic Stimulation: The Importance of Strategy Use

Shane Fresnoza[1,2], Monica Christova[3,4], Sieglinde Purgstaller[1], Margit Jehna[5], Karla Zaar[6], Markus Hoffermann[6], Kariem Mahdy Ali[6], Christof Körner[1,2], Eugen Gallasch[2,3], Gord von Campe[6] and Anja Ischebeck[1,2]*

[1]Institute of Psychology, University of Graz, Graz, Austria, [2]BioTechMed, Graz, Austria, [3]Otto Loewi Research Center, Physiology Section, Medical University of Graz, Graz, Austria, [4]Department of Physiotherapy, University of Applied Sciences FH-Joanneum Graz, Graz, Austria, [5]Department of Radiology, Medical University of Graz, Graz, Austria, [6]Department of Neurosurgery, Medical University of Graz, Graz, Austria

OPEN ACCESS

Edited by:
Peter A. Hall,
University of Waterloo, Canada

Reviewed by:
Vera Moliadze,
University of Kiel, Germany
Matteo Feurra,
National Research University Higher School of Economics, Russia

****Correspondence:***
Shane Fresnoza
shane.fresnoza@uni-graz.at

Specialty section:
This article was submitted to Brain Imaging and Stimulation, a section of the journal Frontiers in Human Neuroscience

Received: *25 September 2019*
Accepted: *16 June 2020*
Published: *16 July 2020*

Citation:
Fresnoza S, Christova M, Purgstaller S, Jehna M, Zaar K, Hoffermann M, Mahdy Ali K, Körner C, Gallasch E, von Campe G and Ischebeck A (2020) Dissociating Arithmetic Operations in the Parietal Cortex Using 1 Hz Repetitive Transcranial Magnetic Stimulation: The Importance of Strategy Use. Front. Hum. Neurosci. 14:271. doi: 10.3389/fnhum.2020.00271

The triple-code model (TCM) of number processing suggests the involvement of distinct parietal cortex areas in arithmetic operations: the bilateral horizontal segment of the intraparietal sulcus (hIPS) for arithmetic operations that require the manipulation of numerical quantities (e.g., subtraction) and the left angular gyrus (AG) for arithmetic operations that require the retrieval of answers from long-term memory (e.g., multiplication). Although neuropsychological, neuroimaging, and brain stimulation studies suggest the dissociation of these operations into distinct parietal cortex areas, the role of strategy (online calculation vs. retrieval) is not yet fully established. In the present study, we further explored the causal involvement of the left AG for multiplication and left hIPS for subtraction using a neuronavigated repetitive transcranial magnetic stimulation (rTMS) paradigm. Stimulation sites were determined based on an fMRI experiment using the same tasks. To account for the effect of strategy, participants were asked whether they used retrieval or calculation for each individual problem. We predicted that the stimulation of the left AG would selectively disrupt the retrieval of the solution to multiplication problems. On the other hand, stimulation of the left hIPS should selectively disrupt subtraction. Our results revealed that left AG stimulation was detrimental to the retrieval and online calculation of solutions for multiplication problems, as well as, the retrieval (but not online calculation) of the solutions to subtraction problems. In contrast, left hIPS stimulation had no detrimental effect on both operations regardless of strategy.

Keywords: number processing, arithmetic, parietal cortex, rTMS, intraparietal sulcus, angular gyrus, strategy

INTRODUCTION

The ability to attend to numbers is innate to some degree in human beings. Discrimination of small numerosities begins during the first weeks of life (Antell and Keating, 1983). By about 5 months after birth, children already attend to the addition or subtraction of one or two items (Wynn, 1992). As, we become acquainted with exact arithmetic during school, our strategies in dealing with different arithmetic problems differ. Direct retrieval of solutions from long-term memory is efficient when solving simple addition and multiplication problems that were taught by rote learning. On the other hand, procedural strategies such as counting ("online calculation") are often used for subtraction, which is often taught by quantity-based counting or other strategies (e.g., inverse addition; Siegler, 1988; Dehaene et al., 2003). Strategy selection, however, depends on several problem-related variables, such as problem size, and individual-related variables, such as working memory span (Imbo and Vandierendonck, 2008). Often, easier problems are solved using retrieval whereas more difficult problems are solved by counting (Zbrodoff and Logan, 2005). A high working-memory span has been linked to the frequent use of retrieval strategies (Imbo and Vandierendonck, 2008).

The triple-code model (TCM) assumes that three different parietal regions are involved in number processing (Dehaene et al., 2003). Based on the findings from neuropsychological, neuroimaging, and brain stimulation studies, the model proposes three distinct and task-specific brain areas in the parietal lobe. The bilateral intraparietal sulcus (IPS) is associated with the core quantity system, the left angular gyrus (AG) is believed to be involved in the verbal processing of numbers, and the posterior superior parietal area in spatial and non-spatial attention (Dehaene and Cohen, 1997; Dehaene et al., 2003). In healthy individuals, arithmetic operations that require online numerical processing such as in simple subtraction and complex (double-digit) addition or multiplication elicited significant unilateral or bilateral IPS activation particularly in its horizontal segment (hIPS; Chochon et al., 1999; Lee, 2000; Menon et al., 2000; Zago et al., 2001; Simon et al., 2002; Delazer et al., 2003, 2005; Ischebeck et al., 2006; Prado et al., 2011; Klein et al., 2013b; De Visscher et al., 2015). The results from these imaging studies seem to support the proposal of the TCM that the hIPS subserve the mental manipulation of numerical quantities (Klein et al., 2013b). This hypothesis was further supported by neuropsychological data showing that pathological lesions of the left and right hIPS caused specific deficits in subtraction with preserved knowledge of rote-learned arithmetic facts (Dehaene and Cohen, 1997; Cohen et al., 2000). Furthermore, findings from non-invasive brain stimulation studies also added evidence that highlighted the importance of the hIPS for arithmetic operations that require online calculation. For example, a virtual lesion-induced on either the right or left hIPS using high frequency repetitive transcranial magnetic stimulation (rTMS) temporarily impaired the participants' ability to solve double-digit addition and subtraction (Göbel et al., 2006; Andres et al., 2011; Montefinese et al., 2017). Cathodal transcranial direct current stimulation (tDCS) over the left posterior parietal cortex also decreased the learning rates for subtraction, whereas anodal tDCS showed an improvement that lasted over 24 h after stimulation (Hauser et al., 2013; Grabner et al., 2015).

Concerning multiplication, the TCM proposes the involvement of the left AG in the retrieval of arithmetic facts which are represented verbally in long-term memory (Dehaene et al., 2003; Klein et al., 2013b). Indeed, significant left AG activation has been reported when healthy individuals encounter low-interfering problems (e.g., simple addition or single-digit multiplication) that are strongly encoded in long-term memory (Stanescu-Cosson et al., 2000; Grabner et al., 2009; Jost et al., 2009; Klein et al., 2013b; De Visscher et al., 2015; Soylu and Newman, 2016). Incorrect or "confusion" equations in which the proposed answer was true for the other operation (e.g., $9 \times 6 = 15$) also elicited increased activation in the left AG because the confusion effect automatically (automatic mapping of the operands of the problems and the associated solutions) activates arithmetic facts in memory (Grabner et al., 2013). Multiplication training also led to increased activation in the left AG due to the shift from quantity-based processing to more automatic retrieval (Ischebeck et al., 2006). Moreover, brain lesions located close to the left AG were shown to induce acalculia for addition, multiplication, and division but with spared subtraction (Lampl et al., 1994; Dehaene and Cohen, 1997; Cohen et al., 2000; Lee, 2000). The findings from invasive and non-invasive brain stimulation studies also support a role of left AG in multiplication. Single-session of anodal tDCS over the right AG elicited bilateral AG activity detected with fMRI for multiplication problems rehearsed during stimulation (Clemens et al., 2013). On the other hand, calculation mapping with 5 Hz rTMS was able to induce a maximum error rate (ER) of 30% in the left AG for a single-digit multiplication task (Maurer et al., 2016). Similarly, direct cortical stimulation (DCS) close to the left AG in patients with tumors in the left parietal area disrupted the performance in single-digit addition, subtraction, and multiplication (Whalen et al., 1997; Duffau et al., 2002; Kurimoto et al., 2006). In a patient with a low-grade glioma in the right temporal cortex, DCS of the right AG also disrupted single-digit subtraction (Yu et al., 2011). Moreover, in some cases, removal of the tumor improved multiplication ability (Kurimoto et al., 2006).

Taken together, the mentioned studies support the direct involvement of the AG in arithmetic operations that need retrieval from memory like multiplication and of the hIPS in arithmetic operations that require online calculation like subtraction. However, findings that challenge this anatomical and functional dissociation of these operations also exist. For instance, a PET study failed to show significant activations on either the left and right AG in the retrieval vs. compute contrast (Zago et al., 2001). Several fMRI studies also showed that retrieval and calculation are not exclusive functions of the left AG and hIPS and a reversal or overlap of function may occur (Fulbright et al., 2000; Delazer et al., 2003; Andres et al., 2011; Arsalidou and Taylor, 2011; Rosenberg-Lee et al., 2011; De Visscher et al., 2015). Common activation patterns distributed in frontoparietal and central regions were also reported when contrasting all arithmetic operations of different complexity. It was suggested that this common activation pattern reflects a basic anatomical substrate of working memory, numerical knowledge, and processing based on finger counting that is derived from a network originally related to finger movements (Fehr et al., 2007). Moreover, findings from lesion and brain stimulation studies added controversial results. Intraoperative DCS during complex 2-digit integer minus 1-digit integer subtraction and single-digit multiplication in both the left AG and left hIPS yielded a similar disruption of processing for both operations in four tumor patients (Pu et al., 2011). Preserved multiplication ability was also reported in a patient with damage to the left AG (van Harskamp et al., 2002). In TMS studies, although low frequency (5 Hz) stimulation of the left and right AG induced 30% and 40% errors in simple multiplication and subtraction, respectively (Maurer et al., 2016), high frequency (10 Hz) rTMS also significantly impaired the performance in complex addition when delivered to the left AG (Göbel et al., 2006). In another study, single-pulse TMS stimulation of the bilateral hIPS disrupted the performance in single-digit addition, while only left hIPS stimulation disrupted single-digit multiplication (Salillas et al., 2012). For tDCS, although bilateral bi-cephalic stimulation of the IPS affects magnitude processing, it does not affect double-digit addition and subtraction task performance (Hauser et al., 2013; Klein et al., 2013a). Moreover, single-session anodal tDCS of the left AG enhanced the RT and decrease the solution rates for large and small addition and subtraction problems, respectively (Rütsche et al., 2015). This overview demonstrates that the complete anatomical and functional dissociation of arithmetic operations in the parietal cortex is far from being clear.

One of the possible reasons for this contradictory pattern of results is the disregard for different strategy use in solving arithmetic problems. Item-by-item strategy use was not fully and correctly accounted for by previous studies. Instead, the two operations, subtraction, and multiplication were commonly used to tap into the brain networks subserving the mental manipulation of numerical quantities and arithmetic fact retrieval, respectively. However, this simple distinction might not be valid for all items. For example, ties (e.g., 3 × 3, 3 + 3) are often solved faster than other problems, which has been attributed to direct memory retrieval (Imbo et al., 2007). It has also been assumed that, in the case of single-digit addition problems, retrieval of arithmetic fact knowledge is used only for rather small problems (e.g., 2 + 3) but not for relatively larger problems (e.g., 8 + 9; Stanescu-Cosson et al., 2000; Klein et al., 2013b). Additionally, retrieval might again be the strategy of choice for multi-digit problems such as 12 + 12 or 20 + 30. This also applies to single-digit multiplication problems because multiplication with zero and small problems are assumed to be solved by rule application and fact retrieval, respectively, and problems with large operands sometimes involve backup strategies when direct retrieval is not sufficient (Jost et al., 2009). Therefore, the majority of the previous studies underestimated the impact of strategy use on an item-by-item basis. Averaging of response latencies across trials that involved different strategies might result in misleading conclusions about how adults solve arithmetic problems (Thevenot et al., 2007). The same critique applies to recent neuroimaging studies. Currently, only one fMRI study (Grabner et al., 2009) has utilized trial-by-trial self-reports to assess strategy usage. So far, no noninvasive brain stimulation study has used this approach to systematically explore the impact of strategy use in subtraction and multiplication. Elucidating the anatomical and functional dissociation of subtraction and multiplication to distinct areas of the parietal cortex will extend our knowledge about the neuronal circuits involved in arithmetic operations. This is useful in understanding the course of disorders like developmental dyscalculia which affects 5–6% of school children, as well as, in formulating interventions for an acquired numerical disability such as in the elderly (Shalev, 2004; Nouchi and Kawashima, 2014).

The present study addressed this issue by using an item-by-item questionnaire to investigate the extent to which the participant's strategy usage affects the anatomical dissociation of multiplication and subtraction. First, we used fMRI to identify the parietal cortex areas recruited during the performance of subtraction and multiplication for each participant. Second, the participants underwent rTMS sessions during which an inhibitory stimulation paradigm (1-Hz rTMS) was applied over three target areas: the left hIPS, left AG, and the vertex as a control site. Participants solved subtraction and multiplication problems before, during, and after stimulation. Immediately after each experimental session, participants were asked to indicate which strategy (online calculation or retrieval) they used to solve each problem using a questionnaire. We predicted that if the left hIPS is engaged in subtraction, the rTMS-induced virtual lesion would increase the solution latency of trials solved by online calculations. On the other hand, if the left AG is engaged in multiplication, the rTMS-induced virtual lesion will increase the solution latency of trials solved by retrieval.

MATERIALS AND METHODS

Participants

The number of participants was determined *a priori* using the statistical software G*Power 3.1.9 (Faul et al., 2007). The estimation indicated that 12 participants would be sufficient in a within-subject repeated measure design (power level of 95% and medium (0.50) effect size). In the study, 16 healthy young volunteers (seven males) with a mean age of 26.25 ± 7.07 (SD) years were recruited. They all had a normal or corrected-to-normal vision and were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). Participants neither had a history of acute or chronic medical or neuropsychiatric diseases and contraindications to TMS such as metallic or electrical implants in the body (Rossi et al., 2009). They received monetary compensation for their participation and gave written informed consent before the experiment. The study protocols complied with the guidelines of the Declaration of Helsinki for human studies and were approved by the ethics committee of the Medical University Graz.

Stimuli and Task

In the fMRI and rTMS experiments, we presented 36 subtraction and 36 multiplication problems. The problems were presented horizontally in white on a black background using Presentation software (Neurobehavioral Systems Inc., Berkely, CA, USA) for the fMRI experiment and Superlab 4.5 software (Cedrus Corporation, San Pedro, CA, USA) for the rTMS experiment (**Figure 1A**). For multiplication problems, one-digit × one-digit multiplications with the numerals from 2 to 9 were selected, including ties. Problems with two different numerals (e.g., 2 × 3) were always presented with the smaller number as the first operator. For subtraction problems, one-digit numerals were subtracted from tens, always requiring a carry operation (e.g., 15–8, but not 15–3).

For the fMRI experiment, the problems were randomized and presented once in a single block (72 trials). The problems were presented together with the solution and a distractor. For multiplication, the distractor was the result of an operant-related multiplication problem. For subtraction, the distractor was either one or two units away from the solution. The distractor was presented on the left side for half of the problems, and on the right side for the other half. After the presentation of the problems and two result alternatives, participants had to press the left or right button to indicate which of the two presented numbers was the solution. During the rTMS experiment, tasks and stimuli were the same as in the fMRI experiment. However, the problems were presented without result alternatives. The participants were asked to mentally solve the problems and speak the solution into a head-mounted microphone connected to a voice-key device. For each rTMS session, the participants solved the 72 problems five times [once before, during, and after (0 min, 30 min, and 60 min) stimulation]. Therefore, one rTMS session had a total number of 360 trials.

Functional Magnetic Resonance Imaging (fMRI)

MRI images were acquired on a 3.0 Tesla whole-body system Siemens Skyra scanner with an echo-planar capable gradient system together with a 20-channel birdcage head coil (Siemens Medical Systems, Erlangen, Germany). For each participant, an anatomical 3-D scan based on a T1-weighted sequence was recorded (TR/TE = 1,650 ms/1.82 ms, matrix = 256 × 256, FOV = 256 mm, 192 sagittal slices, in-plane resolution: 1 mm × 1 mm, slice thickness: 1 mm, 0.5 mm gap). The anatomical scan was followed by functional measurements. For the functional images, a T2*-weighted echo-planar imaging (EPI) sequence was used (TR = 2,000 ms TE = 25 ms, matrix = 74 × 74, FOV = 224 mm, 38 axial slices, in-plane resolution: 3 mm × 3 mm, slice thickness: 3 mm, 0 mm gap) which is sensitive to blood-oxygen-level-dependent (BOLD) contrasts.

Repetitive TMS (rTMS)

The stimulation was performed using a MagPro X100 stimulator with MagOption (MagVenture GmbH, Denmark). Single and repetitive biphasic TMS pulses were delivered using the MCF-B65 and C-B60 figure-of-eight MagVenture coils, respectively. Both coils have a 75 mm diameter on one winding. For stable and precise positioning of the magnetic coil above the areas of interest, the Localite TMS Navigator (Localite GmbH, Sankt Augustin, Germany) system tracks the sensors attached to the coil concerning the adhesive reflectors on the patient's forehead using an infrared tracking device (Polaris Spectra, Northern Digital Inc., Waterloo, ON, Canada). The stimulation intensity was set at 110% of the individual participant's active motor threshold (AMT) determined from the primary motor cortex representation of the right abductor pollicis brevis (APB) muscle using single-pulse TMS. Electromyography (EMG) recordings from the right APB muscle were obtained using surface electrodes with a belly-tendon montage. AMT was defined as the minimum stimulus intensity that elicits a motor-evoked potential (MEP) response of >100 μV (peak-to-peak) during moderate spontaneous background muscle activity (∼10% of the maximum voluntary contraction) in at least five of ten consecutive trials (Rossini et al., 1999). During rTMS stimulation, magnetic pulses were delivered at a frequency of 1 Hz for 15 min (900 pulses; Houdayer et al., 2008). The magnetic coil was held perpendicular to the left hIPS and left AG and was oriented on the central plane at the vertex. All stimulation parameters conformed to the safety guidelines for rTMS (Wassermann, 1998; Rossi et al., 2009).

Experimental Procedure

The study was conducted in a single-blinded, randomized design with an active TMS control condition. Each participant underwent one fMRI and three randomized rTMS sessions (**Figure 1B**). The study always began with the fMRI session. During fMRI, the participants lay supine in the scanner and their head was stabilized with foam paddings. They wore earplugs to protect them from the scanner noise. A computer outside the scanner room controlled the stimulus presentation and scanner triggering (Neurobehavioral Systems Inc., Berkely, CA, USA). The participants viewed the stimulus projected from a monitor at the head end of the scanner on a mirror mounted on top of the head coil. In the fMRI session, each trial started with the presentation of a fixation cross for a jittered duration of 3–11 s (in 500 ms steps, average duration 7 s). Subsequently, the problems with the solutions and distractors appeared for 5 s (**Figure 1A**). Reaction times (RTs) were measured from the onset of the problem presentation until a button press. All 72 trials were presented (without pause) in a single block, leading to a total duration of approximately 13 min. A minimum of 5 days separated the fMRI and the first rTMS session.

All participants underwent three sessions of rTMS stimulation separated by an interval of at least 7 days to avoid carry-over effects. The stimulations were performed in all participants in the middle of the day between 1:00 and 5:00 pm. The stimulation targets (left AG, left hIPS, and vertex) were randomized for each participant. They were not informed about the target locations for each experimental session and the neuronavigation monitor was placed out of their sight to ensure efficient blinding. Vertex stimulation served as the control condition since previous rTMS studies showed that stimulation of this site did not affect number processing (Dormal et al., 2008, 2012; Andres et al., 2011).

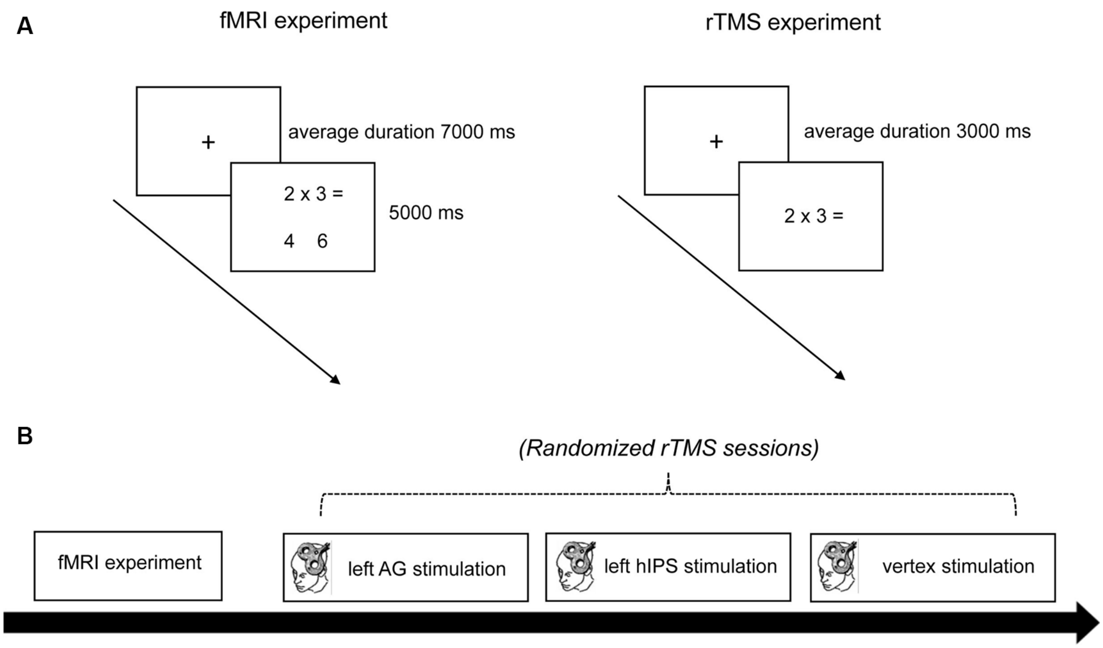


FIGURE 1 | Experimental procedure. **(A)** Trial timing schema of the fMRI and rTMS experiment. In the fMRI experiment, a trial started with the presentation of a fixation cross for an average duration of 7,000 ms. Then the problems (with the solution and distractor) appeared for 5,000 ms. In the rTMS experiments, a trial started with the presentation of a fixation cross for an average duration of 3,000 ms. Then the problems (without the solution and distractor) appeared until the triggering of the voice-key. **(B)** Time course of the experiment. First, anatomical and functional MRI data sets were obtained. Then the participants underwent three sessions of rTMS stimulation (left AG, left hIPS, and vertex) in randomized order. rTMS was applied at 1 Hz for 15 min (900 pulses). fMRI, functional magnetic resonance imaging; rTMS, repetitive transcranial magnetic stimulation; AG, angular gyrus; hIPS, horizontal segment of the intraparietal sulcus.

Additionally, vertex stimulation reproduces the somatosensory effects of parietal stimulations and is considered a better control than other sham stimulation alternatives (Robertson et al., 2003; Dormal et al., 2012). Furthermore, to control for unspecific effects of the stimulation (e.g., motor area), participants performed a grooved pegboard test (PBT) before the first rTMS experimental session and immediately after the last arithmetic task performance (60 min after stimulation) in the third rTMS experimental session (Koch and Rothwell, 2009; Koch et al., 2009; Feurra et al., 2011; Rivera-Urbina et al., 2015).

During rTMS sessions, participants were seated in a comfortable chair with head and armrests. They were informed about the sensations during TMS stimulations and were assured that any calculation impairment would be temporary. The experiment started once all questions were answered. First, we performed the coregistration of the participant's head and the participant's 3D T1-weighted MRI scan. The high-resolution T1 MRI data were loaded into the Localite TMS Navigator System. For the tracking device to locate the individual head and the position of the TMS coil during stimulation, three reflective sphere markers were attached to the patient's forehead and TMS coil. Subsequently, three anatomical landmarks (the nasion, left, and right outer canthus) were marked in the 3D MRI image. Using a digitizing pen that also contained sphere markers, the same anatomic landmarks were marked on the patient's real head. To further improve the co-registration quality, an additional 200 anatomical landmarks were added on the patient's head by tracing the scalp with the digitizing pen. To ensure the goodness of fit (patient's real head and structural MRI), we kept the root mean squared error of the fitting procedure at less than 2.5 mm for all participants. The co-registration created a 3D head model in which the peeling depth could be individually adjusted to visualize the cortical surface.

After the coregistration, the "motor hotspot" or the primary motor cortex representation of the right APB muscle was located using anatomical landmarks (e.g., hand knob at the precentral gyrus). The "motor hotspot" was defined as the cortical location where the lowest stimulator output elicited the biggest MEP amplitudes. EMG electrodes were attached at the right APB muscle in a belly-tendon montage to monitor the MEP amplitudes during stimulation using the built-in EMG device in the stimulator. The participants wore earplugs to shield them from the noise of the stimulator. To confirm the location of the motor hotspot, single-pulse TMS stimulation was applied at a frequency of 0.25 Hz while monitoring MEP amplitudes. The coil was placed tangentially to the scalp at an angle of 45° to the midsagittal plane with the handle pointing laterally and posteriorly generating an anteroposterior current direction in the brain. The participants were asked to briefly and voluntarily contract the APB muscle (∼10% of the maximum voluntary contraction) while TMS was delivered. The stimulation intensity was gradually reduced until the AMT was reached. Participants with an individual AMT beyond 50% of the maximum stimulator output would have been excluded from the experiment (none).

Subsequently, the participant's functional data set was overlaid on the 3D reconstruction. Cortical areas with significant BOLD activations ["fMRI hotspots" or regions of interest (ROI)] were identified and marked.

For the rTMS experiment, each trial started with the presentation of a fixation cross for a variable duration between 2,000 ms and 4,000 ms (in steps of 250 ms, average duration: 3,000 ms). This was followed by the presentation of one of the 72 arithmetic problems (36 subtraction, 36 multiplication). The participants were asked to mentally solve the problem and speak the solution aloud into a head-mounted microphone connected to a voice-key. The problem disappeared on the triggering of the voice-key. After voice-key triggering, the participant's answer was recorded by the experimenter, or a code ("0") for voice-key failure was recorded. The participants solved the arithmetic problems once before, during, and after (0 min, 30 min, and 60 min) stimulation. RT was measured from the onset of the problem presentation until the triggering of the voice key. After each session, participants completed a questionnaire of the 72 arithmetic problems. For each arithmetic problem, they were instructed to tick a box to indicate whether they retrieved the answer from memory or whether they had to calculate. Including the preparation time (20 min), each rTMS experimental session lasted for about 120 min.

DATA ANALYSIS

fMRI Data

Data pre-processing and analysis were performed with SPM12 (Wellcome Department of Cognitive Neurology, London, UK). The first two functional scans of each participant were discarded to allow for signal stabilization. The functional scans were motion-corrected and unwrapped. They were normalized using the MNI functional (EPI) template. Finally, images were spatially smoothed with a Gaussian kernel of 8 mm FWHM. Statistical analyses were performed based on the general linear model as implemented in SPM12. First, a model with two conditions (subtraction/multiplication) was analyzed. To investigate the influence of strategy (calculation or retrieval) participants were asked to complete a questionnaire including all problems before the first rTMS session. For the fMRI analysis, these data were then used to estimate a second model with two conditions (calculation/retrieval). The canonical form of the hemodynamic response function and its first temporal derivative was used for modeling. The motion parameters gained from the motion correction procedure were entered into the model as parameters of no interest. A high-pass filter (cut-off frequency: 1/120 Hz) was applied to remove low-frequency drifts. No global normalization was used. A second level or random-effects analysis was calculated based on the contrast images of the individual subjects (Friston, 1999). The statistical parameter maps were thresholded using an initial uncorrected *p*-value threshold of less than 0.001, reporting only clusters as significant when they had a corrected *p*-value of less than 0.05 on the cluster level. The correction of the *p*-level was based on continuous random field theory as implemented in SPM12 [family-wise error (FWE) corrected].

Behavioral Data (Questionnaire)

Participants had selected either retrieval or calculation as their strategy in the questionnaire, which contained all 72 problems and was administered once after every rTMS session. Only correctly ticked problems were analyzed (3,452 out of 3,456 data points). The percentages for the retrieval strategy were entered into a repeated-measures ANOVA with the operation of the within-subjects factors (subtraction, multiplication) and session (one, two, or three).

Behavioral Data (Reaction Time and Error Rate)

Statistical analyses were conducted separately for the raw RTs and error rates (ERs) during fMRI and rTMS sessions using SPSS software (SPSS 24, IBM Corp., Armonk, NY., USA). In the final analysis, only the RTs from correctly answered and ticked problems were included. Trials for which the RTs were outside of +2 standard deviations and trials with RTs below 300 ms or longer than 5,000 ms (outliers) were excluded. Grouping the RTs according to strategy type produced unbalanced data sets. Therefore, we decided to analyze the RT from the fMRI and three rTMS sessions using a linear mixed-effects model (LMM) because this analysis can accommodate data sets with different numbers of observations per subject (West, 2009). In the models, each participant was specified as a random factor (random intercept model). The RT or ER served as the dependent variable. For the fMRI data sets, a full model included the within-subject factor "operation" (multiplication and subtraction), and "strategy" (calculation and retrieval) as fixed factors. On the other hand, a full model for the rTMS data sets included the within-subject factor "stimulation site" (hIPS, AG, and vertex), "operation" (multiplication and subtraction), "strategy" (calculation and retrieval), and "time" (before, during, and 0, 30 and 60 min after stimulation) as fixed factors.

Normal data distribution (Shapiro–Wilk test) and homogeneity of variance tests (Levene's test) were conducted. To achieve a parsimonious model for the data, we conducted a (forward) stepwise approach by incrementally adding the predictors to a baseline model (Barr et al., 2013). The baseline models only contained the random factor (intercept) to examine the individual variation in the dependent variable without regard to the other predictors (Singer and Willett, 2003). We then added the within-subject factors including their respective interactions to the model one at a time and compared the Akaike Information Criterion (AIC) values that indicate model adequacy. Model over-fitting, particularly for RTs from the rTMS experiment, can be minimized using this method because it penalizes the likelihood function for having too many parameters. Upon the addition of a factor, a decrease or increase in AIC value (>2) indicates model fit improvement or worsening, respectively (Burnham and Anderson, 2002). A maximum likelihood estimation (Compound Symmetry models) was used to estimate the parameters of each model. Additionally, we determined the Akaike weight of each model because the AIC value only compares one model to the next and does not indicate the absolute fit of the model to the data (Burnham and Anderson, 2002). The Akaike weights compare all possible models and determine which model fits the data best for all comparisons. In the final models, we also excluded non-significant factors except when they were involved in significantly higher interactions. Additionally, to test for multicollinearity, we also determined the tolerance and variance inflation factor of the final models. SPSS does not provide effect size values for mixed models, we therefore manually calculated Cohen's d as a measure of effect size. Significant findings from the models were explored using paired t-tests for *post hoc* comparisons (two-tailed, $p < 0.05$, Bonferroni corrected for multiple comparisons). A t-test for dependent measures was used to compare the grooved PBT performance before and after the experiment. A p-value of < 0.05 was considered significant for all statistical analyses. All values are expressed as mean ± standard error of the mean (SEM).

RESULTS

fMRI Data

For the fMRI data, we first contrasted subtraction with multiplication. We found significantly stronger activation for subtraction than for multiplication in the right and left superior parietal lobule, including the IPS and extrastriate visual areas, as well as the right middle frontal gyrus (**Figure 2A**, **Table 1**). No brain area was significantly activated in the reverse contrast. To investigate whether brain activation depended on the strategy used we then contrasted problems whose solutions were reported calculated with problems that were retrieved (**Figure 2B**, **Table 1**). We again found activations in the right and left superior parietal lobule including the IPS, the left precentral gyrus, and left superior frontal gyrus, as well as in the middle cingulate cortex. In the reverse contrast, there was significant activation in the left AG, the right middle temporal gyrus including the AG, the middle cingulate cortex, as well as in the right and left superior frontal gyrus.

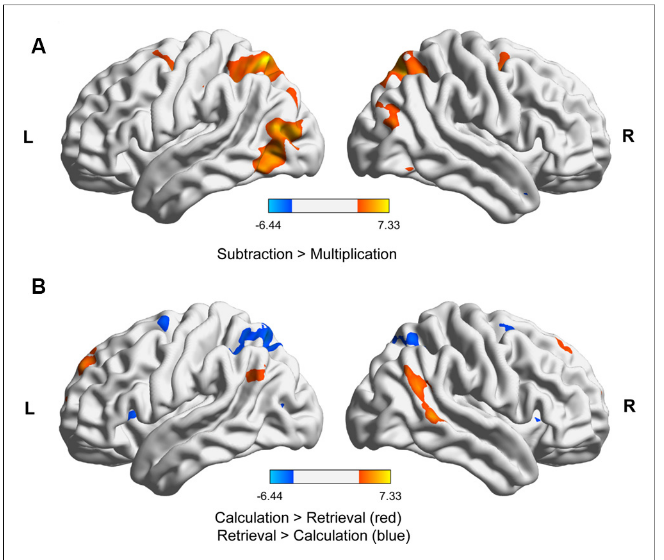


FIGURE 2 | fMRI results for **(A)** subtraction vs. multiplication (both strategies) and **(B)** calculation vs. retrieval (both operations). The activations in the images were thresholded at $p < 0.001$ uncorrected, showing only clusters significant at $p < 0.05$, family-wise error (FWE)-corrected at the cluster level. The images were generated with BrainNet Viewer (Xia et al., 2013).

Behavioral Data (RT and ER) During fMRI

Participants had calculated simple multiplication and subtraction problems during fMRI measurement. Their answers were categorized according to the strategy used for a solution, as given by the questionnaire administered in the first session of the rTMS experiment. For the RT analysis, 1,090 out of 1,152 problems were correctly answered and considered. Additional 16 trials were considered outliers and excluded from the analysis. One additional data point was lost due to missing questionnaire data. Therefore, the final model for the RT contained 1,073 trials or 93.14% of the whole data set. The analysis of the RT revealed that participants were faster with multiplication problems (1.91 s, SD = 0.75 s) than with subtraction problems (2.36 s, SD = 0.85 s), which led to a significant main effect of operation ($F_{(1,556.74)} = 38.94$, $p \leq 0.001$, $d = 0.544$). They were also faster for problems when the solution could be retrieved (1.98 s, SD = 0.39 s) than when the solution had to be calculated (2.26 s, SD = 0.39 s), which is reflected in a significant main effect of strategy ($F_{(1,556.07)} = 15.10$, $p \leq 0.001$, $d = 0.717$). The interaction was not significant ($F_{(1,551.95)} = 0.007$, $p = 0.935$, $d = 0.108$). In the analysis of the ER, no significant effects obtained.

Strategy Questionnaire in the rTMS Experiment

As skilled adults rely on the multiplication tables, we had hypothesized that retrieval was the predominant strategy in multiplication and less so in subtraction. Furthermore, it was expected that participants more often relied on retrieval the more familiar they got with the problems from the first to the third session due to learning. Both expectations were confirmed by our results. Overall, retrieval was more often used for multiplication (70.14%, SD = 11.37%) than for subtraction (27.20%, SD = 11.81%) yielding a main effect of operation ($F_{(1,15)} = 44.31$, $p < 0.001$, $\eta_p^2 = 0.75$). Furthermore, participants used the retrieval strategy more in the later sessions (session 1: 42.97%, SD = 7.91%, session 2: 50.04%, SD = 8.87%, and session 3: 52.60%, SD = 9.11%, main effect session ($F_{(2,30)} = 5.70$, $p = 0.008$, $\eta_p^2 = 0.28$). The interaction was not significant. This indicates that familiarity had a similar effect on strategy use for both operations.

TMS Parameters and Impact on Motor Function

During the TMS sessions, all participants tolerated the single and repetitive TMS stimulations well. The mean stimulation intensities (hIPS: 42.75 ± 1.65% MSO, AG 40.93 ± 1.59% MSO, vertex: 43.12 ± 1.83% MSO) did not significantly differ between the sessions. There were no reports of headaches, dizziness, or nausea. In four participants, we noticed some episodes of difficulty verbally naming the solution for multiplication problems during AG stimulation but they were able to finish the experiments. The results of the grooved PBT indicated that

TABLE 1 | Brain areas activated for subtraction vs. multiplication (both strategies) and retrieval vs. calculation targets (both operations).

Hemisphere		x	y	z	K	Z
	Subtraction > Multiplication					
Left	Superior parietal lobule	−26	−66	60	4,519	4.95
Right	Superior parietal lobule	16	−72	62	1,785	4.45
Right	Middle frontal gyrus	24	−2	48	1,082	4.28
	Multiplication > Subtraction					
	ns.					
	Retrieval > Calculation					
Right	Middle cingulate cortex	6	−40	36	778	4.71
Right	Middle temporal gyrus	62	−46	4	598	4.41
Left	Superior frontal gyrus	−14	56	38	225	4.05
Right	Superior frontal gyrus	16	44	52	160	3.95
Left	Angular gyrus	−50	−60	32	141	3.60
	Calculation > Retrieval					
Right	Middle cingulate cortex	10	22	40	710	4.39
Left	Superior parietal lobule	−16	−64	60	1,441	4.26
Right	Superior parietal lobule	26	−62	54	310	3.81
Left	Precentral gyrus	−50	6	26	191	3.81
Left	Superior frontal gyrus	−28	0	70	170	3.75

Statistical parameter maps were thresholded with an initial threshold of p < 0.001 uncorrected, reporting only clusters that survived an family-wise error (FWE)-corrected p-value < 0.05. Coordinates are reported as given by SPM12 (MNI space). k = cluster size, Z = Z value for the maximally activated voxel of the cluster.

our stimulation protocol had no significant impact on motor function (remote effect) of the right (before: 58.25 ± 1.65 s, after: 57.63 ± 1.64 s, $t_{(15)} = 0.379$, $p = 0.710$) and left hand (before: 60.31 ± 1.88 s, after: 62.69 ± 2.13 s, $t_{(15)} = -1.488$, $p = 0.158$). PBT performance were also comparable between the participants who received vertex ($n = 6$; right hand: 57.66 ± 3.67 s, left hand: 63.16 + 4.96 s), left hIPS ($n = 5$; right hand: 58.20 + 2.72 s, left hand: 63.00 + 2.50 s), and left AG ($n = 5$; right hand: 57.00 + 2.16 s, left hand: 61.80 + 3.30 s) stimulation on their last experimental session (all $p \geq 0.05$).

Behavioral Data (RT and ER) in the rTMS Experiment

For the RT, we decided to interpret a full model because all the main effects were highly significant (**Table 2**), the addition of each variable and their interactions improved the model based on the AIC values, and a model containing the 4-way interactions did come out best 100% of the time based on the Akaike weights (**Supplementary Table S1**). For the three rTMS sessions, we included 93.32% (16,127 trials out of 17,280) of the RT data in the final analysis. The raw data entered in the final model were normally distributed after log transformation (Shapiro–Wilk test) and the variances were equal (Levene's test; all $p > 0.05$). Multicollinearity was not a concern in the final model since the tolerance range and variance inflation factors were 0.863–1.00 and 1.000–1.159, respectively. The RT data from the three rTMS sessions are presented in **Figures 3A,B**. These data are normalized to their respective baseline measures to remove baseline differences between the sessions. The results of the analysis (performed on the raw data, not normalized to baseline data) revealed significant differences in RT before and after rTMS stimulation of the three target areas (significant main effect of time: $F_{(4,16111.02)} = 5.07$, $p \leq 0.001$, $d = 0.078$; significant main effect of stimulation site: $F_{(2,16111.27)} = 23.11$, $p \leq 0.001$, $d = 0.175$; and significant time and stimulation site interactions: $F_{(8,16111.25)} = 4.40$, $p \leq 0.001$, $d = 0.214$; **Figures 3A,B**). The *post hoc* comparisons for the factor time showed that participants were significantly faster in solving arithmetic problems 60 min after stimulation compared to before ($p = 0.001$) and during ($p = 0.030$) stimulation. They were specifically faster in solving arithmetic problems when the left hIPS was stimulated compared to the vertex ($p \leq 0.001$) and AG ($p \leq 0.001$; pairwise comparisons, Bonferroni corrected; **Figures 3A,B**). The analysis also showed that the participants were slower in solving multiplication than subtraction problems (significant main effect of operation: $F_{(1,16116.85)} = 112.73$, $p \leq 0.001$, $d = 0.350$) and slower in retrieving the answer compared to calculating it particularly 60 min after stimulation (significant main effect of strategy: $F_{(1,16126.54)} = 434.45$, $p \leq 0.001$, $d = 0.775$; and significant time and strategy interactions: $F_{(4,16111.01)} = 5.25$, $p \leq 0.001$, $d = 0.159$). Concerning the site-specific effect, rTMS stimulation of the left AG slowed down the online calculation or retrieval process in both operations (significant operation and strategy interactions: $F_{(1,16122.48)} = 45.40$, $p \leq 0.001$, $d = 0.350$; significant strategy and stimulation site interactions: $F_{(2,16111.99)} = 6.26$, $p = 0.002$, $d = 0.185$). In contrast, similar to the vertex stimulation, rTMS of the hIPS did not inhibit the online calculation and retrieval of answers to both multiplication and subtraction problems.

Regarding the ERs, the participants exhibited very low ERs before the stimulation (4.44% in multiplication and 4.46% in subtraction). The ER further decreased after stimulation in all conditions as indicated by the significant main effect of time only ($F_{(4,358.79)} = 3.66$, $p = 0.006$, $d = 0.420$; **Supplementary Table S2**).

DISCUSSION

The present study aimed at elucidating the anatomical and functional dissociation of subtraction and multiplication into

TABLE 2 | Results of the linear mixed model (LMM) performed on the reaction times from the repetitive transcranial magnetic stimulation (rTMS) experiment.

	Numerator *df*	Denominator *df*	*F*-value	*p*-value	Cohen's *D*
Time	4	16,111.02	5.07	<0.001*	0.078
Operation	1	16,116.85	112.73	<0.001*	0.350
Strategy	1	16,126.54	434.45	<0.001*	0.775
Stimulation site	2	16,111.27	23.11	<0.001*	0.175
Time × operation	4	16,111.00	0.79	0.534	0.096
Time × strategy	4	16,111.01	5.25	<0.001*	0.159
Time × stimulation site	8	16,111.25	4.40	<0.001*	0.214
Operation × strategy	1	16,122.48	45.40	<0.001*	0.350
Operation × stimulation site	2	16,111.21	1.78	0.164	0.175
Strategy × stimulation site	2	16,111.99	6.26	0.002*	0.185
Time × operation × strategy	4	16,111.00	0.75	0.555	0.229
Time × operation × stimulation site	8	16,111.00	0.22	0.987	0.153
Time × strategy × stimulation site	8	16,111.02	1.71	0.090	0.176
Operation × strategy × stimulation site	2	16,111.82	2.52	0.081	0.189
Time × operation × strategy × stimulation site	8	16,111.00	1.26	0.262	0.184

For the LMM (random intercept model), each participant was treated as a random factor. The within-subjects factor stimulation site (hIPS, AG, and vertex), operation (multiplication and subtraction), strategy (calculation and retrieval), and time (before, during, and 0, 30 and 60 min after stimulation) were treated as fixed factors. Asterisks indicate significant results ($p < 0.05$). df = Degrees of freedom.

distinct parietal cortex areas namely the left hIPS and left AG, respectively. First, we identified brain areas with significant activation during the performance of subtraction and multiplication using fMRI. Second, these brain areas were stimulated using rTMS. We were particularly interested in the impact of the participant's strategy of choice on the dissociation of these two operations. Therefore, we used a strategy questionnaire to have first-hand knowledge of how the participants solved subtraction and multiplication problems. The strategy questionnaires alone revealed that multiplication compared to subtraction problems were more often solved using a retrieval strategy. fMRI data analysis revealed significant recruitment of the left AG during retrieval (more than online calculation), even though we did not observe a significant increase in activity at the left AG during multiplication (compared to subtraction). Conversely, we observed stronger activation in the bilateral hIPS during subtraction (more than in multiplication) and online calculation (more than for retrieval). Our fMRI findings corroborate the results of previous imaging studies highlighting the role of the left AG in multiplication problems that require retrieval strategy and the bilateral hIPS for subtraction problems that require online calculation strategy (Delazer et al., 2003, 2005; Ischebeck et al., 2006; Grabner et al., 2009). Additionally, our imaging results also showed significant activations of the prefrontal, frontal, and cingulate cortices during calculation and retrieval. Activations of these areas indicate their involvement in the strategy selection network in number processing that requires working memory, strategic organization during encoding, decision making, and response selection (Taillan et al., 2015). In the rTMS sessions, our results showed that left AG stimulation was detrimental to the retrieval and online calculation of solutions for multiplication problems, as well as, the retrieval (but not online calculation) of the solutions to subtraction problems. In contrast, left hIPS stimulation had no detrimental effect on both operations regardless of strategy.

RTMS Stimulation of the Left AG

The stimulation of the left AG resulted in marked RT slowing in multiplication (more than subtraction) problems which indicate an impairment in our participants' ability to perform this arithmetic operation. Our result provides further support for the assumption that the left AG is crucial in solving arithmetic problems that are typically solved by the retrieval of the solution from verbal long-term memory (Cohen et al., 2000; Dehaene et al., 2003; Seghier, 2013; Andin et al., 2015). However, when we analyzed the RT based on strategy, the results were contrary to our expectations because the impairment was smaller in magnitude for multiplication problems solved using retrieval compared to the online calculation. Retrieval was only impaired during and immediately after the stimulation, while online calculation was impaired until 30 min after stimulation. Furthermore, we also observed impairment in the retrieval of the solutions to subtraction problems, particularly during the stimulation. Therefore, our results suggest that the left AG plays a role in the retrieval of the solution from memory for both multiplication and subtraction problems. Our results further suggest that the left AG is also responsible for the online calculation of solutions to multiplication problems.

The impairment in retrieving the solution to multiplication problems was expected because retrieval of overlearned multiplication facts (e.g., 2 × 3) is supported by language-relevant areas that include the left AG (Dehaene et al., 2003). This is demonstrated among adult individuals with deficits in phonological processing, such as those with developmental dyslexia who have prominent difficulties in multiplication due to poor retrieval of arithmetic facts (De Smedt and Boets, 2010). This is because arithmetic facts are represented verbally in long-term memory, allowing such problems to be solved by arithmetic fact retrieval (Klein et al., 2013b). In our study, the close functional interplay of arithmetic fact retrieval and language processing was demonstrated in four participants who exhibited difficulties to verbalize the result for multiplication problems during left AG stimulation. The interference probably

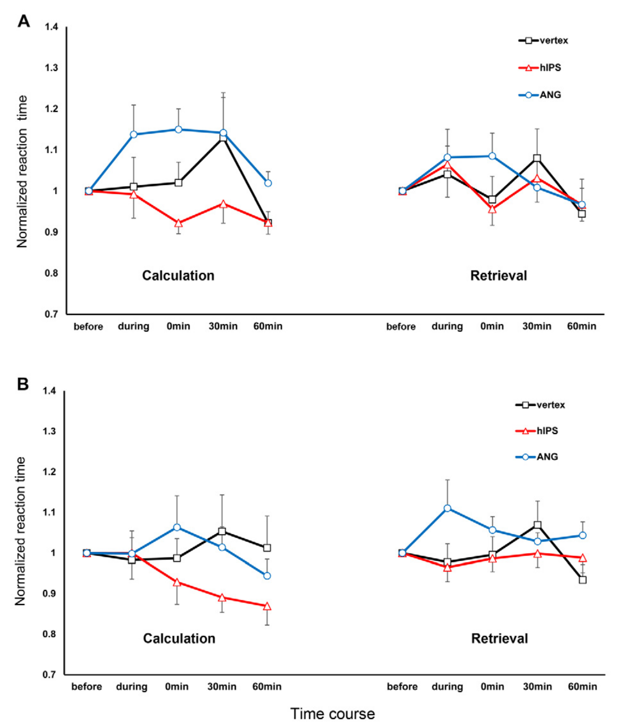


FIGURE 3 | Response times for multiplication and subtraction problems, depending on time and site of stimulation, separated for operation and strategy. The *x*-axis displays the time points. The *y*-axis represents the reaction times (RTs) normalized concerning the respective RT before the stimulation (RT during and after stimulation/RT before stimulation). **(A)** Multiplication: The stimulation of the AG inhibited the online calculation and retrieval of solutions for multiplication problems. The stimulation of the hIPS and vertex did not inhibit online calculation and retrieval. **(B)** Subtraction: stimulation of the AG inhibited the retrieval of solutions to subtraction problems during stimulation. Stimulation of the hIPS had no detrimental effect on online calculation or retrieval. Error bars represent the standard error of the mean. hIPS, the horizontal segment of the left intraparietal sulcus; AG, angular gyrus.

involved a genuine impairment of arithmetic fact retrieval because the production task (verbal response) put stronger demands on the retrieval of the correct answer from memory than the solution selection task used for the fMRI experiment (Dehaene et al., 1999; Andres et al., 2011). We can rule out the possibility that the speech interruptions were due to motor impairment because the effect was specific to the stimulation of left AG during multiplication. Also, the stimulation had no impact on grooved PBT performance. The impairment in the retrieval of the solution to subtraction problems could be anticipated because some subtraction problems (e.g., 12–6) may be stored in verbal long-term memory as well. Indeed, impairment in single-digit addition, subtraction, and multiplication can be elicited by directly stimulating the cortical areas close to the left AG during tumor surgery (Whalen et al., 1997; Duffau et al., 2002; Kurimoto et al., 2006). In our arithmetic task, even though we did not use single-digit operands, retrieval might have been the strategy of choice for some subtraction problems such as in 16–8 because 16 is double the amount of 8. To conclude, the impairment in the retrieval of solution for both operations indicates that the interference in automatic fact retrieval is due to the rTMS-induced tonic suppression of neuronal activity in the left AG (Ridding and Ziemann, 2010).

One might ask if arithmetic operations solved by retrieval involve the left AG (Dehaene et al., 2003), why is the retrieval not fully disrupted by the stimulation? The AG has strong functional and anatomical connectivity with the hippocampal system and the frontal areas mainly *via* the middle longitudinal fascicle (ventral pathway). This is different from the IPS, which is connected by the superior longitudinal fascicle (dorsal pathway) with frontal areas for magnitude-related processes, as revealed by probabilistic fiber tracking (Klein et al., 2013b). Additionally, dorsal fiber tracts like the cortical cingulate route (via retrosplenial cortex) that provide an indirect pathway for hippocampal interactions with prefrontal cortex were also described to subserve arithmetic fact retrieval (Uddin et al., 2010; Klein et al., 2013b; Bubb et al., 2017). The retrosplenial cortex was reported to be involved in the recognition of familiar objects and procedures, as well as autobiographical memory. This function is related to the retrieval of familiar arithmetic facts from memory (Vann et al., 2009; Sestieri et al., 2010, 2013; Klein et al., 2013b). Possibly the left AG stimulation might not have been sufficient to completely inhibit the retrieval process since other brain areas (e.g., retrosplenial cortex) subserving memory retrieval were less affected by the inhibitory effect of the stimulation. This is because the inhibitory effect of 1-Hz rTMS is mainly localized in the cortex being stimulated which in our case was the left AG. As shown in *in vivo* electrophysiological studies in the human motor cortex, 1-Hz rTMS only suppresses the late I-waves that depend on the excitability of motor cortico-cortical circuits (Di Lazzaro et al., 2003, 2010; Cirillo and Perez, 2015). Indeed, anodal tDCS of the AG also failed to affect multiplication performance despite significant BOLD activation in the retrosplenial cortex (Clemens et al., 2013). This could explain the short duration of retrieval impairments (only lasted immediately after stimulation), as well as, the low ER (5.4%) we and another rTMS study (30%) observed after left AG stimulation (Maurer et al., 2016). This reasoning might also explain why a lesion of the left AG is neither a sufficient nor a necessary condition to observe a deficit in multiplication (van Harskamp et al., 2002, 2005).

For the online calculation of the solution, stimulation of the left AG elicited robust RT slowing that lasted for 30 min in multiplication, while in subtraction RT slowing was only observed immediately after the stimulation. The impairment in the online calculation was also unexpected because arithmetic problems that require quantity manipulations were thought to be processed in the hIPS (Dehaene et al., 2003). It is therefore unclear, why left AG stimulation markedly disrupted online calculation of solution to multiplication problems. In theory, the strategy-of-choice for simple multiplication problems is retrieval. However, when retrieval fails, for instance when faced with more complex operations such as multi-digit multiplication or interference due to stimulation, a participant may adaptively use another strategy (e.g., online calculation) to produce a response. For instance, whenever direct fact retrieval for an arithmetic problem fails, bilateral intraparietal areas may be involved in semantic re-coding of the problem, recruiting magnitude information of the numbers involved (Dehaene, 1995; Klein et al., 2013b). This might have been the scenario in our participants because retrieval was impaired during and immediately after the stimulation of the left AG. However, if online calculation involves the decomposition of the arithmetic problem into smaller facts (e.g., $14 - 8 = 14 - 4 = 10 - 4 = 6$), impaired retrieval of these smaller facts from verbal long-term memory will in turn negatively affect the efficiency of procedural strategy (De Smedt and Boets, 2010). Therefore, we argue that the impairment in the online calculation of answers could be secondary to the impairment in retrieval. As reflected by our imaging results, the strong activations in the cingulate, motor and frontal cortices might reflect not only the increased conflict during the fact-retrieval processes but also higher demands for controlling and coordinating multiple processing steps when a problem cannot be solved by direct retrieval (Jost et al., 2009). Additionally, the use of online calculation would be a costly strategy because this puts higher demands on verbal working memory, which might lead to slower performance in solving multiplication problems (Hecht, 2002; De Smedt and Boets, 2010). This could explain why the performance in double-digit additions that were probably solved using online calculation (more than retrieval) was also disrupted by bilateral DCS and rTMS stimulation of the AG (Roux et al., 2003; Göbel et al., 2006; Montefinese et al., 2017). On the other hand, subtraction problems solved by online calculations were not profoundly affected by the stimulation of the left AG because this strategy was thought to be carried out by the hIPS (Dehaene et al., 2003). As shown by our results, the RTs for subtraction problems solved by online calculations were not markedly prolonged by the stimulation of the left AG, as well as the vertex. In contrast, operations (double-digit addition and subtraction) that require online calculation were significantly impaired by rTMS stimulation of the left or bilateral hIPS (Göbel et al., 2006; Montefinese et al., 2017).

RTMS Stimulation of the Left hIPS

The results from the stimulation of the left hIPS were also unexpected because we initially predicted that left hIPS stimulation would impair our participants' ability to solve arithmetic problems that require genuine quantity manipulations such as subtraction (Dehaene et al., 2003). Instead, we did not observe any detrimental effects such as RTs slowing or increased ER in subtraction as well as in multiplication problems during and after left hIPS stimulation. Nevertheless, the effect of stimulation on RTs was strategy-dependent: retrieval was not affected whereas online calculation was improved by the stimulation in both operations. Retrieval was comparable in both sham and left hIPS stimulation conditions indicating that the left hIPS had no or only a minimal role in the retrieval of solutions from memory in subtraction and multiplication problems. Moreover, our behavioral finding was consistent with our imaging results because we did not observe significant hIPS activation during retrieval. Therefore, we argue that retrieval was not affected by the stimulation of the left hIPS because this strategy does not entirely depend on it. On the other hand, RTs for problems solved by online calculation decreased after left hIPS stimulation indicating an improvement in our participants' ability to solve both operations using this strategy. Our imaging results also showed significant activations of the bilateral hIPS during the online calculation. This was consistent with the reported recruitment of brain areas involved in numerical quantity processing when participants were solving untrained (calculated) more than trained (mostly retrieved) subtraction and multiplication problems (Simon et al., 2002; Ischebeck et al., 2006). In contrast, our behavioral results did not corroborate the findings of previous rTMS studies that showed performance disruption in arithmetic operations (e.g., double-digit addition and subtraction) that need online calculation (Göbel et al., 2006; Montefinese et al., 2017). The performance improvement could not be due to a learning effect because it was specific for problems solved using an online calculation. Here, we may ask, why would an inhibitory rTMS stimulation paradigm applied to the left hIPS improve online calculation? For subtraction, one possible reason is that we did not stimulate and therefore inhibit the right hIPS. According to previous studies, subtraction-related areas are also predominantly localized in the right hIPS (Cohen et al., 2000; Andres et al., 2011; Maurer et al., 2016). This argument is in good accordance with the recent results from Montefinese et al. (2017) that highlighted the role of the right hIPS, as well as, the right ventral segment of IPS (vIPS) in solving complex arithmetic operations. In their study, bilateral hIPS and vIPS high frequency rTMS stimulation disrupted double-digit addition and subtraction. They argued that the stimulation disrupted online calculation because during complex arithmetic problem solving our reliance on visuospatial strategies, a suggested function of the right IPS, increases (Montefinese et al., 2017). In theory, the complexity of our subtraction problems (e.g., the requirement to conduct a "carry" procedure) may have facilitated the recruitment of the right IPS and engage visuospatial strategies as shown by the bilateral hIPS activation during the online calculation. Therefore, the recruitment of the right hIPS and the use of visuospatial strategies might have facilitated task performance because this strategy not only enhances numerosity processing and length categorization but also the processing of serial position information on the spatially oriented mental number line in mental arithmetic (Dormal et al., 2012; Knops and Willmes, 2014; Montefinese et al., 2017). Indeed, impairment not only in numerical but also in spatial bisection tasks was reported in patients with a lesion in the right parietal cortex (Zorzi et al., 2002; Cappelletti et al., 2007). Our results also showed that online calculation improvement was more robust in subtraction than multiplication problems. Here, we suggest that subtraction was less affected by inhibitory stimulation because subtraction-related areas of the cortex are known for being robust toward brain lesions or aphasia, in contrast to multiplication- or division-related cortical areas

(Lampl et al., 1994; Pesenti et al., 1994; Maurer et al., 2016). Lastly, we also suggest the same arguments to explain the performance improvement in multiplication problems solved by online calculation. A study highlighted the similar role of the right IPS in multiplication by showing that single-pulse rTMS stimulation of IPS in either hemisphere (compared to control sites) led to increased RTs in addition and multiplication. They suggest that computational efficiency is not specifically dependent on left hemisphere regions and that efficiency in multiplication is dependent on the right vIPS considered to be critical for motion representation and automatization (Salillas et al., 2012).

CONCLUSION

The present findings emphasized the presence of two distinct cortical networks that are modulated by the strategy and not by the arithmetic operation *per se*. For instance, we have shown that the integrity of the left AG is required in performing retrieval and online calculation strategy in multiplication, but only for the retrieval strategy in subtraction. On the other hand, the results from the stimulation of the left hIPS may indirectly suggest that the integrity of the right hIPS was sufficient to perform both operations, particularly when using the online calculation strategy. However, we would like to emphasize that great care must be taken in correlating our results with previous rTMS studies because none of them took into account the strategy used by the participants. The same principle must be applied in interpreting the correlation between our results and the findings from brain imaging studies in healthy participants, as well as, electrophysiological and neuropsychological studies in patients. This is because neuroimaging can elucidate brain areas involved in a certain task but it does not allow any causal interpretation, that is, it cannot be deduced from neuroimaging alone which areas are indeed essential for calculation. Studies done on tumor patients (mostly single-case studies) should also be interpreted cautiously since slow-growing tumors can shift the calculation-related areas and affect other parietal areas that are involved in arithmetic operations. Overall, the present findings addressed some of the disparities from previous studies. Most importantly, our findings can be a basis for developing therapeutic interventions aimed at reducing the effects of developmental dyscalculia or acquired numerical disability (Lepage and Théoret, 2010). It was already shown that increasing the excitability of the right and left parietal cortex in healthy adult participants using tDCS improved numerical ability (e.g., greater learning rates for subtraction) that lasted for 24 h up to 6 months after stimulation (Cohen Kadosh et al., 2010; Grabner et al., 2015). However, this effect was not replicated in a pilot study on two adults with developmental dyscalculia (Iuculano and Cohen Kadosh, 2014). Therefore, further investigations are warranted particularly those that focus on the strategy that these individuals are often using when solving numerical problems.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee of the Medical University Graz. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SF, CK, EG, GC, and AI designed the study. SF, SP, MJ, and KZ conducted fMRI experiments. SF, MC, SP, MH, KM, and KZ conducted the rTMS experiments. SF and AI analyzed the data and wrote the manuscript. SF, MC, SP, MJ, KZ, MH, KM, EG, CK, GC, and AI reviewed and approved the final version of the manuscript.

FUNDING

The authors acknowledge the financial support by the University of Graz.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00271/full#supplementary-material.

REFERENCES

Andin, J., Fransson, P., Rönnberg, J., and Rudner, M. (2015). Phonology and arithmetic in the language-calculation network. *Brain Lang.* 143, 97–105. doi: 10.1016/j.bandl.2015.02.004

Andres, M., Pelgrims, B., Michaux, N., Olivier, E., and Pesenti, M. (2011). Role of distinct parietal areas in arithmetic: an fMRI-guided TMS study. *NeuroImage* 54, 3048–3056. doi: 10.1016/j.neuroimage.2010.11.009

Antell, S. E., and Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Dev.* 54, 695–701. doi: 10.2307/1130057

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

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001

Bubb, E. J., Kinnavane, L., and Aggleton, J. P. (2017). Hippocampal - diencephalic - cingulate networks for memory and emotion: an anatomical guide. *Brain Neurosci. Adv.* 1:2398212817723443. doi: 10.1177/2398212817723443

Burnham, K., and Anderson, D. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* New York, NY: Springer-Verlag.

Cappelletti, M., Freeman, E. D., and Cipolotti, L. (2007). The middle house or the middle floor: bisecting horizontal and vertical mental number lines in neglect. *Neuropsychologia* 45, 2989–3000. doi: 10.1016/j.neuropsychologia.2007.05.014

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.* 11, 617–630. doi: 10.1162/089892999563689

Cirillo, J., and Perez, M. A. (2015). Subcortical contribution to late TMS-induced I-waves in intact humans. *Front. Integr. Neurosci.* 9:38. doi: 10.3389/fnint.2015.00038

Clemens, B., Jung, S., Zvyagintsev, M., Domahs, F., and Willmes, K. (2013). Modulating arithmetic fact retrieval: a single-blind, sham-controlled tDCS study with repeated fMRI measurements. *Neuropsychologia* 51, 1279–1286. doi: 10.1016/j.neuropsychologia.2013.03.023

Cohen Kadosh, R., Soskic, S., Iuculano, T., Kanai, R., and Walsh, V. (2010). Modulating neuronal activity produces specific and long-lasting changes in numerical competence. *Curr. Biol.* 20, 2016–2020. doi: 10.1016/j.cub.2010.10.007

Cohen, L., Dehaene, S., Chochon, F., Lehéricy, S., and Naccache, L. (2000). Language and calculation within the parietal lobe: a combined cognitive, anatomical and fMRI study. *Neuropsychologia* 38, 1426–1440. doi: 10.1016/s0028-3932(00)00038-5

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 Visscher, A., Berens, S. C., Keidel, J. L., Noël, M.-P., and Bird, C. M. (2015). The interference effect in arithmetic fact solving: an fMRI study. *NeuroImage* 116, 92–101. doi: 10.1016/j.neuroimage.2015.04.063

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

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., Spelke, E., Pinel, P., Stanescu, R., and Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* 284, 970–974. doi: 10.1126/science.284.5416.970

Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Lochy, A., Trieb, T., et al. (2003). Learning complex arithmetic—an fMRI study. *Cogn. Brain Res.* 18, 76–88. doi: 10.1016/j.cogbrainres.2003.09.005

Delazer, M., Ischebeck, A., Domahs, F., Zamarian, L., Koppelstaetter, F., Siedentopf, C. M., et al. (2005). Learning by strategies and learning by drill—evidence from an fMRI study. *NeuroImage* 25, 838–849. doi: 10.1016/j.neuroimage.2004.12.009

Di Lazzaro, V., Oliviero, A., Mazzone, P., Pilato, F., Saturno, E., Dileone, M., et al. (2003). Chapter 12 generation of I waves in the human: spinal recordings *Supplements to Clinical Neurophysiology*, eds W. Paulus, F. Tergau, M. A. Nitsche, J. G. Rothwell, U. Ziemann and M. Hallett (Amsterdam, AE: Elsevier), 143–152.

Di Lazzaro, V., Profice, P., Pilato, F., Dileone, M., Oliviero, A., and Ziemann, U. (2010). The effects of motor cortex rTMS on corticospinal descending activity. *Clin. Neurophysiol.* 121, 464–473. doi: 10.1016/j.clinph.2009.11.007

Dormal, V., Andres, M., and Pesenti, M. (2008). Dissociation of numerosity and duration processing in the left intraparietal sulcus: a transcranial magnetic stimulation study. *Cortex* 44, 462–469. doi: 10.1016/j.cortex.2007.08.011

Dormal, V., Andres, M., and Pesenti, M. (2012). Contribution of the right intraparietal sulcus to numerosity and length processing: an fMRI-guided TMS study. *Cortex* 48, 623–629. doi: 10.1016/j.cortex.2011.05.019

Duffau, H., Denvil, D., Lopes, M., Gasparini, F., Cohen, L., Capelle, L., et al. (2002). Intraoperative mapping of the cortical areas involved in multiplication and subtraction: an electrostimulation study in a patient with a left parietal glioma. *J. Neurol. Neurosurg. Psychiatry* 73, 733–738. doi: 10.1136/jnnp.73.6.733

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/bf03193146

Fehr, T., Code, C., and Herrmann, M. (2007). Common brain regions underlying different arithmetic operations as revealed by conjunct fMRI-BOLD activation. *Brain Res.* 1172, 93–102. doi: 10.1016/j.brainres.2007.07.043

Feurra, M., Bianco, G., Polizzotto, N. R., Innocenti, I., Rossi, A., and Rossi, S. (2011). Cortico-cortical connectivity between right parietal and bilateral primary motor cortices during imagined and observed actions: a combined TMS/tDCS study. *Front. Neural Circuits* 5:10. doi: 10.3389/fncir.2011.00010

Friston, K. J. (1999). Schizophrenia and the disconnection hypothesis. *Acta Psychiatr. Scand.* 395, 68–79. doi: 10.1111/j.1600-0447.1999.tb05985.x

Fulbright, R. K., Molfese, D. L., Stevens, A. A., Skudlarski, P., Lacadie, C. M., and Gore, J. C. (2000). Cerebral activation during multiplication: a functional MR imaging study of number processing. *Am. J. Neuroradiol.* 21, 1048–1054. Available online at: http://www.ajnr.org/content/21/6/1048.abstract.

Göbel, S. M., Rushworth, M. F., and Walsh, V. (2006). Inferior parietal RTMS affects performance in an addition task. *Cortex* 42, 774–781. doi: 10.1016/s0010-9452(08)70416-7

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., and Ebner, F. (2013). The function of the left angular gyrus in mental arithmetic: evidence from the associative confusion effect. *Hum. Brain Mapp.* 34, 1013–1024. doi: 10.1002/hbm.21489

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., and Neuper, C. (2009). To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. *Neuropsychologia* 47, 604–608. doi: 10.1016/j.neuropsychologia.2008.10.013

Grabner, R. H., Rütsche, B., Ruff, C. C., and Hauser, T. U. (2015). Transcranial direct current stimulation of the posterior parietal cortex modulates arithmetic learning. *Eur. J. Neurosci.* 42, 1667–1674. doi: 10.1111/ejn.12947

Hauser, T. U., Rotzer, S., Grabner, R. H., Mérillat, S., and Jäncke, L. (2013). Enhancing performance in numerical magnitude processing and mental arithmetic using transcranial Direct Current Stimulation (tDCS). *Front. Hum. Neurosci.* 7:244. doi: 10.3389/fnhum.2013.00244

Hecht, S. A. (2002). Counting on working memory in simple arithmetic when counting is used for problem solving. *Mem. Cognit.* 30, 447–455. doi: 10.3758/bf03194945

Houdayer, E., Degardin, A., Cassim, F., Bocquillon, P., Derambure, P., and Devanne, H. (2008). The effects of low- and high-frequency repetitive TMS on the input/output properties of the human corticospinal pathway. *Exp. Brain Res.* 187, 207–217. doi: 10.1007/s00221-008-1294-z

Imbo, I., and Vandierendonck, A. (2008). Effects of problem size, operation and working-memory span on simple-arithmetic strategies: differences between children and adults? *Psychol. Res.* 72, 331–346. doi: 10.1007/s00426-007-0112-8

Imbo, I., Vandierendonck, A., and Rosseel, Y. (2007). The influence of problem features and individual differences on strategic performance in simple arithmetic. *Mem. Cognit.* 35, 454–463. doi: 10.3758/bf03193285

Ischebeck, A., Zamarian, L., Siedentopf, C., Koppelstätter, F., Benke, T., Felber, S., et al. (2006). How specifically do we learn? Imaging the learning of multiplication and subtraction. *NeuroImage* 30, 1365–1375. doi: 10.1016/j.neuroimage.2005.11.016

Iuculano, T., and Cohen Kadosh, R. (2014). Preliminary evidence for performance enhancement following parietal lobe stimulation in Developmental Dyscalculia. *Front. Hum. Neurosci.* 8:38. doi: 10.3389/fnhum.2014.00038

Jost, K., Khader, P., Burke, M., Bien, S., and Rösler, F. (2009). Dissociating the solution processes of small, large, and zero multiplications by means of fMRI. *NeuroImage* 46, 308–318. doi: 10.1016/j.neuroimage.2009.01.044

Klein, E., Mann, A., Huber, S., Bloechle, J., Willmes, K., Karim, A. A., et al. (2013a). Bilateral bi-cephalic tdcs with two active electrodes of the same polarity modulates bilateral cognitive processes differentially. *PLoS One* 8:e71607. doi: 10.1371/journal.pone.0071607

Klein, E., Moeller, K., Glauche, V., Weiller, C., and Willmes, K. (2013b). Processing pathways in mental arithmetic—evidence from probabilistic fiber tracking. *PLoS One* 8:e55455. doi: 10.1371/journal.pone.0055455

Knops, A., and Willmes, K. (2014). Numerical ordering and symbolic arithmetic share frontal and parietal circuits in the right hemisphere. *NeuroImage* 84, 786–795. doi: 10.1016/j.neuroimage.2013.09.037

Koch, G., and Rothwell, J. C. (2009). TMS investigations into the task-dependent functional interplay between human posterior parietal and motor cortex. *Behav. Brain Res.* 202, 147–152. doi: 10.1016/j.bbr.2009.03.023

Koch, G., Ruge, D., Cheeran, B., Fernandez Del Olmo, M., Pecchioli, C., Marconi, B., et al. (2009). TMS activation of interhemispheric pathways between the posterior parietal cortex and the contralateral motor cortex. *J. Physiol.* 587, 4281–4292. doi: 10.1113/jphysiol.2009.174086

Kurimoto, M., Asahi, T., Shibata, T., Takahashi, C., Nagai, S., Hayashi, N., et al. (2006). Safe removal of glioblastoma near the angular gyrus by awake surgery preserving calculation ability—case reportandmdash. *Neurol. Med. Chir.* 46, 46–50. doi: 10.2176/nmc.46.46

Lampl, Y., Eshel, Y., Gilad, R., and Sarova-Pinhas, I. (1994). Selective acalculia with sparing of the subtraction process in a patient with left parietotemporal hemorrhage. *Neurology* 44, 1759–1761. doi: 10.1212/wnl.44.9.1759

Lee, K.-M. (2000). Cortical areas differentially involved in multiplication and subtraction: a functional magnetic resonance imaging study and correlation with a case of selective acalculia. *Ann. Neurol.* 48, 657–661. doi: 10.1002/1531-8249(200010)48:4<657::aid-ana13>3.0.co;2-k

Lepage, J.-F., and Théoret, H. (2010). Numerical processing: stimulating numbers. *Curr. Biol.* 20, R975–R977. doi: 10.1016/j.cub.2010.10.017

Maurer, S., Tanigawa, N., Sollmann, N., Hauck, T., Ille, S., Boeckh-Behrens, T., et al. (2016). Non-invasive mapping of calculation function by repetitive navigated transcranial magnetic stimulation. *Brain Struct. Funct.* 221, 3927–3947. doi: 10.1007/s00429-015-1136-2

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

Montefinese, M., Turco, C., Piccione, F., and Semenza, C. (2017). Causal role of the posterior parietal cortex for two-digit mental subtraction and addition: a repetitive TMS study. *NeuroImage* 155, 72–81. doi: 10.1016/j.neuroimage.2017.04.058

Nouchi, R., and Kawashima, R. (2014). Improving cognitive function from children to old age: a systematic review of recent smart ageing intervention studies. *Adv. Neurosci.* 2014:15. doi: 10.1155/2014/235479

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Pesenti, M., Seron, X., and Van Der Linden, M. (1994). Selective impairment as evidence for mental organisation of arithmetical facts: BB, a case of preserved subtraction? *Cortex* 30, 661–671. doi: 10.1016/s0010-9452(13)80242-0

Prado, J., Mutreja, R., Zhang, H., Mehta, R., Desroches, A. S., Minas, J. E., et al. (2011). Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. *Hum. Brain Mapp.* 32, 1932–1947. doi: 10.1002/hbm.21159

Pu, S., Li, Y.-N., Wu, C.-X., Wang, Y.-Z., Zhou, X.-L., and Jiang, T. (2011). Cortical areas involved in numerical processing: an intraoperative electrostimulation study. *Stereotact. Funct. Neurosurg.* 89, 42–47. doi: 10.1159/000321186

Ridding, M. C., and Ziemann, U. (2010). Determinants of the induction of cortical plasticity by non-invasive brain stimulation in healthy subjects. *J. Physiol.* 588, 2291–2304. doi: 10.1113/jphysiol.2010.190314

Rivera-Urbina, G. N., Batsikadze, G., Molero-Chamizo, A., Paulus, W., Kuo, M.-F., and Nitsche, M. A. (2015). Parietal transcranial direct current stimulation modulates primary motor cortex excitability. *Eur. J. Neurosci.* 41, 845–855. doi: 10.1111/ejn.12840

Robertson, E. M., Théoret, H., and Pascual-Leone, A. (2003). Studies in cognition: the problems solved and created by transcranial magnetic stimulation. *J. Cogn. Neurosci.* 15, 948–960. doi: 10.1162/089892903770007344

Rosenberg-Lee, M., Chang, T. T., Young, C. B., Wu, S., and Menon, V. (2011). Functional dissociations between four basic arithmetic operations in the human posterior parietal cortex: a cytoarchitectonic mapping study. *Neuropsychologia* 49, 2592–2608. doi: 10.1016/j.neuropsychologia.2011.04.035

Rossi, S., Hallett, M., Rossini, M. P., and Pascual-Leone, A. (2009). *Safety, Ethical Considerations, and Application Guidelines for the Use of Transcranial Magnetic Stimulation in Clinical Practice and Research.* Oxford: Elsevier.

Rossini, P. M., Berardelli, A., Deuschl, G., Hallett, M., Maertens de Noordhout, A. M., Paulus, W., et al. (1999). Applications of magnetic cortical stimulation. The international federation of clinical neurophysiology. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 52, 171–185.

Roux, F.-E., Boetto, S., Sacko, O., Chollet, F., and Trémoulet, M. (2003). Writing, calculating and finger recognition in the region of the angular gyrus: a cortical stimulation study of Gerstmann syndrome. *J. Neurosurg.* 99, 716–727. doi: 10.3171/jns.2003.99.4.0716

Rütsche, B., Hauser, T. U., Jäncke, L., and Grabner, R. H. (2015). When problem size matters: differential effects of brain stimulation on arithmetic problem solving and neural oscillations. *PLoS One* 10:e0120665. doi: 10.1371/journal.pone.0120665

Salillas, E., Semenza, C., Basso, D., Vecchi, T., and Siegal, M. (2012). Single pulse TMS induced disruption to right and left parietal cortex on addition and multiplication. *NeuroImage* 59, 3159–3165. doi: 10.1016/j.neuroimage.2011.10.093

Seghier, M. L. (2013). The angular gyrus: multiple functions and multiple subdivisions. *Neuroscientist* 19, 43–61. doi: 10.1177/1073858412440596

Sestieri, C., Capotosto, P., Tosoni, A., Luca Romani, G., and Corbetta, M. (2013). Interference with episodic memory retrieval following transcranial stimulation of the inferior but not the superior parietal lobule. *Neuropsychologia* 51, 900–906. doi: 10.1016/j.neuropsychologia.2013.01.023

Sestieri, C., Shulman, G. L., and Corbetta, M. (2010). Attention to memory and the environment: functional specialization and dynamic competition in human posterior parietal cortex. *J. Neurosci.* 30, 8445–8456. doi: 10.1523/JNEUROSCI.4719-09.2010

Shalev, R. S. (2004). Developmental dyscalculia. *J. Child Neurol.* 19, 765–771. doi: 10.1177/08830738040190100601

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

Simon, O., Mangin, J.-F., Cohen, L., Le Bihan, D., and Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron* 33, 475–487. doi: 10.1016/s0896-6273(02)00575-5

Singer, J. D., and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence.* New York, NY: Oxford University Press.

Soylu, F., and Newman, S. (2016). Anatomically ordered tapping interferes more with one-digit addition than two-digit addition: a dual-task fMRI study. *Cogn. Process.* 17, 67–77. doi: 10.1007/s10339-015-0737-2

Stanescu-Cosson, R., Pinel, P., van de Moortele, P.-F., Le Bihan, D., Cohen, L., and Dehaene, S. (2000). Understanding dissociations in dyscalculia. A brain imaging study impact number size cerebral networks exact approx calculation. *Brain* 123, 2240–2255. doi: 10.1093/brain/123.11.2240

Taillan, J., Ardiale, E., Anton, J.-L., Nazarian, B., Félician, O., and Lemaire, P. (2015). Processes in arithmetic strategy selection: a fMRI study. *Front. Psychol.* 6:61. doi: 10.3389/fpsyg.2015.00061

Thevenot, C., Fanget, M., and Fayol, M. (2007). Retrieval or nonretrieval strategies in mental arithmetic? An operand recognition paradigm. *Mem. Cognit.* 35, 1344–1352. doi: 10.3758/bf03193606

Uddin, L. Q., Supekar, K., Amin, H., Rykhlevskaia, E., Nguyen, D. A., Greicius, M. D., et al. (2010). Dissociable connectivity within human angular gyrus and intraparietal sulcus: evidence from functional and structural connectivity. *Cereb. Cortex* 20, 2636–2646. doi: 10.1093/cercor/bhq011

van Harskamp, N. J., Rudge, P., and Cipolotti, L. (2002). Are multiplication facts implemented by the left supramarginal and angular gyri? *Neuropsychologia* 40, 1786–1793. doi: 10.1016/s0028-3932(02)00036-2

van Harskamp, N. J., Rudge, P., and Cipolotti, L. (2005). Does the left inferior parietal lobule contribute to multiplication facts? *Cortex* 41, 742–752. doi: 10.1016/s0010-9452(08)70293-4

Vann, S. D., Aggleton, J. P., and Maguire, E. A. (2009). What does the retrosplenial cortex do? *Nat. Rev. Neurosci.* 10, 792–802. doi: 10.1038/nrn2733

Wassermann, E. M. (1998). Risk and safety of repetitive transcranial magnetic stimulation: report and suggested guidelines from the International Workshop on the Safety of Repetitive Transcranial Magnetic Stimulation, June 5-7, 1996. *Electroencephalogr. Clin. Neurophysiol.* 108, 1–16. doi: 10.1016/s0168-5597(97)00096-8

West, B. T. (2009). Analyzing Longitudinal Data With the Linear Mixed Models Procedure in SPSS. *Eval. Health Prof.* 32, 207–228.

Whalen, J., McCloskey, M., Lesser, R. P., and Gordon, B. (1997). Localizing arithmetic processes in the brain: evidence from a transient deficit during cortical stimulation. *J. Cogn. Neurosci.* 9, 409–417. doi: 10.1162/jocn.1997.9.3.409

Wynn, K. (1992). Addition and subtraction by human infants. *Nature* 358, 749–750. doi: 10.1038/358749a0

Xia, M., Wang, J., and He, Y. (2013). BrainNet viewer: a network visualization tool for human brain connectomics. *PLoS One* 8:e68910. doi: 10.1371/journal.pone.0068910

Yu, X., Chen, C., Pu, S., Wu, C., Li, Y., Jiang, T., et al. (2011). Dissociation of subtraction and multiplication in the right parietal cortex: evidence from intraoperative cortical electrostimulation. *Neuropsychologia* 49, 2889–2895. doi: 10.1016/j.neuropsychologia.2011.06.015

Zago, L., Pesenti, M., Mellet, E., Crivello, F., Mazoyer, B., and Tzourio-Mazoyer, N. (2001). Neural correlates of simple and complex mental calculation. *NeuroImage* 13, 314–327. doi: 10.1006/nimg.2000.0697

Zbrodoff, N. J., and Logan, G. D. (2005). "What everyone finds: the problem-size effect," in *Handbook of Mathematical Cognition*, ed. J. I. D. Campbell (New York, NY: Psychology Press), 331–345.

Zorzi, M., Priftis, K., and Umiltà, C. (2002). Neglect disrupts the mental number line. *Nature* 417, 138–139. doi: 10.1038/417138a

Conflict of Interest: 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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 17 July 2020
doi: 10.3389/fnhum.2020.00272

Persistent Differences in Brain Structure in Developmental Dyscalculia: A Longitudinal Morphometry Study

Ursina McCaskey[1,2], Michael von Aster[1,2,3,4], Ruth O'Gorman[1,2,5] and Karin Kucian[1,2,4]*

[1]Center for MR-Research, University Children's Hospital, Zurich, Switzerland, [2]Children's Research Center, University Children's Hospital, Zurich, Switzerland, [3]Clinic for Child and Adolescent Psychiatry, German Red Cross Hospitals, Berlin, Germany, [4]Neuroscience Center Zurich, University of Zurich and ETH Zurich, Zurich, Switzerland, [5]Zurich Center for Integrative Human Physiology, University of Zurich, Zurich, Switzerland

OPEN ACCESS

Edited by:
Korbinian Moeller, Loughborough University, United Kingdom

Reviewed by:
Patrick Müller, German Center for Neurodegenerative Diseases (DZNE), Germany
Ann Dowker, University of Oxford, United Kingdom

***Correspondence:**
Ursina McCaskey
ursina.mccaskey@kispi.uzh.ch

Specialty section:
This article was submitted to Cognitive Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *21 December 2019*
Accepted: *16 June 2020*
Published: *17 July 2020*

Citation:
McCaskey U, von Aster M, O'Gorman R and Kucian K (2020) Persistent Differences in Brain Structure in Developmental Dyscalculia: A Longitudinal Morphometry Study. Front. Hum. Neurosci. 14:272. doi: 10.3389/fnhum.2020.00272

Developmental dyscalculia (DD) is a learning disability affecting the acquisition of numerical-arithmetical skills. Affected people show persistent deficits in number processing, which are associated with aberrant brain activation and structure. Reduced gray matter has been reported in DD for the parietal cortex including the intraparietal sulcus (IPS), but also the frontal and occipito-temporal cortex. Furthermore, dyscalculics show white matter differences for instance in the inferior (ILF) and superior longitudinal fasciculus (SLF). However, the longitudinal development of these structural differences is unknown. Therefore, our goal was to investigate the developmental trajectory of gray and white matter in children with and without DD. In this longitudinal study, neuropsychological measures and T1-weighted structural images were collected twice with an interval of 4 years from 13 children with DD (8.2–10.4 years) and 10 typically developing (TD) children (8.0–10.4 years). Voxel-wise estimation of gray and white matter volumes was assessed using voxel-based morphometry for longitudinal data. The present findings reveal for the first time that DD children show persistently reduced gray and white matter volumes over development. Reduced gray matter was found in the bilateral inferior parietal lobes including the IPS, supramarginal gyri, left precuneus, cuneus, right superior occipital gyrus, bilateral inferior and middle temporal gyri, and insula. White matter volumes were reduced in the bilateral ILF and SLF, inferior fronto-occipital fasciculus (IFOF), corticospinal tracts, and right anterior thalamic radiation (ATR). Behaviorally, children with DD performed significantly worse in various numerical tasks at baseline and follow-up, corroborating persistent deficits in number processing. The present results are in line with the literature showing that children with DD have reduced gray and white matter volumes in the numerical network. Our study further sheds light on the trajectory of brain development, revealing that these known structural differences in the long association fibers and the adjacent regions of the temporal- and frontoparietal cortex persist in dyscalculic children from childhood into adolescence. In conclusion, our results underscore that DD is a persistent learning disorder accompanied by deficits in number processing and reduced gray and white matter volumes in number related brain areas.

Keywords: developmental dyscalculia, longitudinal, gray matter, white matter, children, development, voxel-based morphometry

INTRODUCTION

Numbers and mathematics are omnipresent in our daily lives and their mastery is crucial to function effectively in our society. Poor numeracy skills, therefore, pose a serious burden for persons affected. Developmental dyscalculia (DD) is a learning disorder characterized by significant and persisting difficulties in learning academic skills related to mathematics or arithmetic. The difficulties are not due to a disorder of intellectual development, sensory impairment, mental or neurological disorders, or inadequate instruction (WHO, 2018). DD affects about 3–7% of the school children (Wyschkon et al., 2009; Butterworth et al., 2011) and has been shown to have a persisting character (Shalev et al., 2005; Geary et al., 2013; McCaskey et al., 2018). Studies in children with DD reveal impairments in numerical magnitude processing and difficulties in the retrieval of arithmetical facts from memory, but also in visuospatial memory or inhibition (Geary, 1993; Landerl et al., 2004; De Smedt et al., 2013; Szucs et al., 2013). These deficiencies have been linked to abnormalities in brain function and structure. When processing numbers and performing arithmetic, a large neural network is involved including posterior parietal (intraparietal sulcus (IPS), angular gyrus, supramarginal gyrus), prefrontal, occipito-temporal and hippocampal areas. Children with DD show aberrant activation of the numerical neural network (Price et al., 2007; Davis et al., 2009; Kucian et al., 2011; Ashkenazi et al., 2012) and abnormalities in different measures of brain structure (e.g., fractional anisotropy, cortical thickness, cortical surface area, gray and white brain volumes; Rykhlevskaia et al., 2009; Kucian et al., 2013; Ranpura et al., 2013).

Hitherto, few cross-sectional studies investigated structural differences in white and gray matter volumes in children with DD compared to typically developing (TD) peers (Rotzer et al., 2008; Rykhlevskaia et al., 2009; Ranpura et al., 2013). Generally, all studies report reduced gray matter volumes in dyscalculics in the IPS and the inferior and superior parietal lobes (Rotzer et al., 2008; Rykhlevskaia et al., 2009; Ranpura et al., 2013). These regions have been linked to number processing and mathematical problem-solving in several studies (Dehaene et al., 2003; for a meta-analysis see Arsalidou et al., 2018). Furthermore, decreased volumes are detected in regions of the frontal lobe such as the anterior cingulum and the inferior and middle frontal gyrus (Rotzer et al., 2008), known to be involved in working memory, attention and goal-directed behavior (Arsalidou et al., 2018), and in occipital regions such as the cuneus/precuneus, lateral occipital cortex, lingual and fusiform gyrus (Rykhlevskaia et al., 2009), which process visual numerical information. Finally, less gray matter volume is also found in the entorhinal cortex, the parahippocampal gyrus and the hippocampus (Rykhlevskaia et al., 2009; Ranpura et al., 2013), which is suggested to play an important role in the formation of long-term memory for arithmetical facts (Menon, 2016). In contrast to the results in children, the findings in studies with dyscalculic adults are less clear. A recent study found no differences in volumetric or surface characteristics of gray matter in dyscalculic adults with and without comorbid dyslexia compared to a control group (Moreau et al., 2019). Likewise, Cappelletti and Price (2013) did not find differences in gray matter volume in a whole-brain analysis in adults with DD but could show that the right parietal area had significantly reduced gray matter volume in a region of interest (ROI) analysis. However, a study investigating the structural correlates of mathematical expertise revealed higher gray matter volume in the right superior parietal lobe, but lower gray matter volume in the right IPS in professional mathematicians compared to non-mathematicians (Popescu et al., 2019).

Differences between children with and without DD are not only found for gray matter structures since white matter volume has also been reported to be reduced in children with DD. Less white matter is observed in temporoparietal regions (right inferior parietal lobe, temporal pole, transverse temporal lobe) and the left frontal lobe (Rotzer et al., 2008; Rykhlevskaia et al., 2009; Ranpura et al., 2013). These regions are part of prominent white matter tracts. The inferior (ILF) and superior longitudinal fasciculus (SLF) have been suggested to be particularly important for numerical processing, as they may be involved in frontoparietal communication and visual processing of numerical or mathematical problems (van Eimeren et al., 2008; Tsang et al., 2009; Matejko and Ansari, 2015). Further white matter differences are found in regions that correspond to the inferior fronto-occipital fasciculus (IFOF), forceps major, corticospinal tract (CST), and the anterior thalamic radiation (ATR). Interestingly, fractional anisotropy (FA), a measure for white matter integrity, is reduced in the SFL, ILF, IFOF and the caudal forceps major in children with DD (Rykhlevskaia et al., 2009; Kucian et al., 2013; for a review see Matejko and Ansari, 2015). Note that no studies to date have reported increased gray or white matter volumes for persons with DD compared to peers without DD (Rotzer et al., 2008; Rykhlevskaia et al., 2009; Cappelletti and Price, 2013; Ranpura et al., 2013; Moreau et al., 2019).

Currently, there is one study that investigated children between 8 and 14 years and aimed to describe how the differences between dyscalculic and control children vary during cortical development (Ranpura et al., 2013). Relative to TD children, gray matter volume in dyscalculics increases with age in the left dorsolateral prefrontal cortex and the right superior occipital lobe, but decreases slightly in the left primary motor cortex. White matter development of children with DD showed notable delays relative to the control group. Whilst TD children showed an age-related increase in frontal and parietal areas, the white matter volumes of DD children remained stable over time (Ranpura et al., 2013).

The gray and white matter regions and their relationship with numerical or arithmetical skills were also investigated in TD children. Gray matter volume is positively correlated with arithmetic scores or performance gains specifically in the left IPS and angular gyrus (Li et al., 2013; Supekar et al., 2013; Evans et al., 2015). Price et al. (2016) showed that gray matter volume in the left IPS measured in first grade predicted the math score at the end of the second grade. This result is not confined to the parietal cortex. Increased gray matter volume in the frontal (dorsal and ventral prefrontal cortices, IFG), occipito-temporal (cuneus, fusiform gyrus) and in the hippocampus also relates to better math performance (Li et al., 2013; Supekar et al., 2013; Evans et al., 2015; Wilkey et al., 2018). Moreover, associations between brain volume abnormalities and math performance have been reported for the gray matter volume of parietal regions in other populations prone to math difficulties (prematurely born children, very low birth weight, Turner syndrome; Isaacs et al., 2001; Starke et al., 2013; Zhao et al., 2013).

To summarize, gray matter volume in various regions of the frontoparietal network, but specifically in the IPS—which is known as a key area for number processing—has been associated with better performance in numerical processing and arithmetical skills. In line with that, children with DD show reduced gray matter volumes in parietal, but also frontal, occipito-temporal and hippocampal areas. Furthermore, white matter differences have been reported in the main white matter tracts connecting the parietal with the frontal and the temporal cortex in subjects with DD. However, to the author's best knowledge, there is no study investigating the development of these structural differences in children with DD using a longitudinal study design. Therefore, the present work aims to elucidate the developmental trajectory of gray and white brain matter volume in children with and without DD from childhood to adolescence.

Based on previous literature, we expect to replicate the known group differences. Reduced gray matter volumes in various regions of the frontoparietal numerical network are expected in children with DD compared to an age-matched control group with normal math performance (Rotzer et al., 2008; Rykhlevskaia et al., 2009; Ranpura et al., 2013). Furthermore, we hypothesize that less white matter volume will be present in the main tracts connecting the parietal with the frontal and the temporal lobe for dyscalculic children (Rotzer et al., 2008; Rykhlevskaia et al., 2009). Second, we will examine the general developmental effects of white and gray matter substance. We anticipate an increase in white matter and possibly a decrease in the gray matter over the examined time from childhood to early adolescence. Longitudinal studies focusing on the structural development from childhood to adulthood showed that gray matter volume increases in the first 10 years of life followed by a decrease in the next decades (Mills et al., 2016). However, the peak of the gray matter volume varies though between studies and also brain regions (Giedd et al., 1999; Gogtay and Thompson, 2010; Groeschel et al., 2010; Mills et al., 2016). Findings regarding white matter volumes revealed a constant increase from childhood until young adulthood (Giedd et al., 1999; Groeschel et al., 2010; Mills et al., 2016). Finally, we are interested in the longitudinal trajectory of the group differences. According to the results of Ranpura et al. (2013), we expect a relative increase in gray matter volume and no change in white matter volume over time in children with DD compared to TD peers. In contrast, the findings in adults point towards a normalization of the gray matter structure over time, as no or only little differences were found in dyscalculic adults. Based on this literature we, therefore, expect to find divergent developmental trajectories in dyscalculic compared to controls.

MATERIALS AND METHODS

Participants and Procedure

A total of 35 (23 DD, 12 TD) children between 8 and 11 years were recruited into this longitudinal study, of which 27 took part in a previous study (Kucian et al., 2011). This longitudinal study included structural and functional MRI data (for the results of fMRI data please see McCaskey et al., 2018). Children were evaluated by neuropsychological tests and MRI at baseline and returned after 4.2 (SD = 0.46) years for a follow-up measurement. We approached the subjects of our study through the school setting or School psychological Services (DD subjects). The children visited us twice at the Center for MR-Research of the University Children's Hospital Zurich. On both occasions, they first completed a neuropsychological session and then underwent the MRI measurement.

Inclusion criteria for all children were no history of neurological or psychiatric disorders and an IQ $\geq$ 85, measured by the third edition of the WISC (Tewes et al., 1999; Similarities, Block Design, Vocabulary, Picture Arrangement). Furthermore, DD children had to score below the 10th percentile in the total score or three subtests of a standardized numerical test battery (ZAREKI-R) at baseline. These criteria are in line with the diagnostic criteria for DD of the ICD-11 (6A03.2 Developmental learning disorder with impairment in mathematics; WHO, 2018). TD children had to perform above the cut-off of value in the test batteries for numerical abilities at baseline and follow-up (10th percentile in the ZAREKI-R and 67 points in the BASIS-MATH 4–8, respectively). Following these criteria, six DD children and one TD child were excluded from the study. For the MRI analysis, two children were excluded due to missing imaging data at baseline or follow-up and three for image quality reasons, resulting in 13 DD and 10 TD complete data sets for the whole study. Groups did not differ in age, gender, handedness, and pubertal status determined by the Edinburgh Handedness Inventory (Oldfield, 1971) and an adapted version of the Self-administered Rating Scale for Pubertal Development (Carskadon and Acebo, 1993).

Parental consent and child assent were obtained at the beginning of the study. The study was approved by the

Ethics committee of Zurich, Switzerland based on guidelines from the World Medical Association's Declaration of Helsinki (WMA, 2013).

Neuropsychological Testing

During the neuropsychological assessment, we acquired numerical abilities, general cognitive abilities as well as measures for the most common comorbid disorders such as developmental dyslexia, attention deficit and hyperactivity disorder, and working memory deficits.

Numerical Abilities

At baseline, numerical abilities were assessed using the revised version of the Neuropsychological Test Battery for Number Processing and Calculation in Children (ZAREKI-R; von Aster et al., 2006). The Zareki-R is a multidimensional test, measuring basic numerical skills as well as calculation, and widely used for the diagnosis of DD in the German-speaking area (see **Supplementary Material** for detailed information about the subtests). Based on this test battery children scoring below the 10th percentile in the total score or three subtests were identified with DD.

Also, the Arithmetic subtest of the Wechsler Intelligence Scale for Children (WISC-III; Tewes et al., 1999) was performed. In this subtest, children had to solve story problems of increasing difficulty within a set time limit (reported test values are IQ scores).

At follow-up, the numerical achievement was assessed with the test for Basic Diagnosis in Mathematics Education for Grades 4-8 (BASIS-MATH 4-8; Moser Opitz et al., 2010). The Basis-Math is a criterion-based test battery measuring various arithmetical abilities such as counting, decimal system, and calculation. Criteria for numerical deficiencies are met if the performance is under a threshold value of 67 points (maximum score of 83 points). This is interpreted as not reaching mastery of basic mathematical concepts (see **Supplementary Material** for detailed information).

The curriculum-based subtest Quantity Comparison of the Cognitive Abilities Test (KFT 4-12+R; Heller and Perleth, 2000) was performed to assess the arithmetic performance at a peer level. In this subtest, subjects had 10 min to solve as many quantity comparisons as possible of increasing difficulty (reported test values are T scores).

The spatial representation of numbers was measured using a computerized number line task adopted from Kucian et al. (2011; for a detailed description see McCaskey et al., 2018). Children had to indicate by mouse-click the position of 20 Arabic digits on a number line with the labeled endpoints 0 and 100. Accuracy was measured by calculating the percentage distance from the marked to the correct position of the given number (reported measures are raw values).

Children also solved 40 basic arithmetic problems (20 addition and 20 subtraction) in the number range 0–1,000 (for a detailed description see McCaskey et al., 2018). Each problem was presented visually on the computer screen and solutions were given *via* the keyboard. The number of correctly solved items was quantified (reported test values are raw scores, maximum value 20).

Domain General Cognitive Abilities

At baseline, intelligence was measured with the third edition of the WISC (Similarities, Block Design, Vocabulary, Picture Arrangement; Tewes et al., 1999). At follow-up, the fourth edition of the WISC was used (Similarities, Block Design, Matrix Reasoning; Petermann and Petermann, 2007). **Table 1** shows the estimated general IQ.

Working Memory

Visuospatial working memory was measured with the Block-Suppression-Test (Beblo et al., 2004). The task required subjects to reproduce every second block of a previously presented sequence on a board with nine cubes. The sequences had a length of 3–9 cubes. Three items per sequence were presented. The longest sequence which was reproduced correctly twice was quantified (reported test values are raw scores, maximum value 9). Verbal working memory was measured with the subtest Digit Span of the WISC-IV (Petermann and Petermann, 2007). In this task, subjects had to repeat an auditorily presented sequence of numerals forward or backward. The sequences had a length of 2–9 numerals (reported test values are IQ scores).

Attention

Levels of attention and inhibition were measured using the subtests Alertness and Go-Nogo of the computerized Test battery for Attentional Performance (TAP; Zimmermann and Fimm, 1993). In the Alertness subtest, subjects had to react as quickly as possible when the target stimulus "x" appeared (intrinsic alertness). Half of the trials were preceded by an acoustic cue stimulus (phasic alertness). The test has four runs and a total of 80 target items. For each subject, the percentile rank of the median RT was quantified (reported test values are percentile ranks). In the Go-Nogo subtest, subjects had to react as quickly as possible to a target stimulus ("x," go condition), but inhibit reactions on a second presented stimulus ("+," nogo condition). The test has a total of 40 items (20 go and 20 nogo items). For each subject, the percentile rank of the median RT was quantified (reported test values are percentile ranks).

Reading Abilities

The 1-Minute-Reading-Task from the Salzburg Reading and Orthography Test (SLRT-II; Moll and Landerl, 2010) assessing word and pseudoword reading fluency was used to estimate the reading performance. Two sheets of paper with either 156 words or 156 pseudowords of increasing length and difficulty were presented. Subjects had 1 min per sheet to read as many words as possible. The amount of correctly read items was quantified (reported test values are percentile ranks). Because of lacking test norms in grades 7 and 8, we interpolated the norms from the test manual (grade 6) and Kronschnabel et al. (2013; grade 9).

Behavioral Data Analysis

Behavioral data were statistically analyzed with SPSS (Version 22). To account for the difficulties regarding the performance of statistical tests of normality in small samples, we performed nonparametric tests (Mann–Whitney U Test) to assess group

TABLE 1 | Demographic characteristics and scores on numerical abilities, domain-general cognitive abilities, working memory, attention, and reading.

Behavioral measure	DD		TD		Test-statistic	*p*	*r*
	N	*M (SD)*	*N*	*M (SD)*			
Baseline assessment							
Age	13	9.5 (0.7)	10	9.2 (0.8)	56.0[a]	0.605	0.13
Gender m/f	13	3/10	10	5/5	1.81[b]	0.221	
Handedness l/a/r	13	1/4/8	10	1/3/6	0.04[b]	0.999	
Numerical abilities							
DD diagnosis *(ZAREKI-R)*	13	6.3 (5.0)	10	75.6 (19.5)	0.00[a]	0.000***	0.84
Arithmetic *(WISC-III)*	12	90.4 (9.6)	10	105.5 (12.8)	21.5[a]	0.008**	0.55
Domain general cognitive abilities							
Estimated IQ *(WISC-III)*	13	99.8 (5.8)	10	111.6 (6.9)	11.5[a]	0.000***	0.69
Working memory							
Visuo-spatial *(BST)*	11	2.7 (1.5)	10	3.6 (1.0)	37.5[a]	0.195	0.31
Follow-up assessment							
Age	13	13.5 (0.9)	10	13.6 (0.8)	64.0[a]	0.976	0.01
Puberty Score	13	2.8 (0.7)	10	2.6 (0.8)	57.0[a]	0.636	0.10
Numerical abilities							
DD diagnosis *(BASIS-MATH 4-8)*	13	49.8 (9.1)	10	75.3 (4.2)	0.00[a]	0.000***	0.84
Quantity Comparison *(KFT 4-12+R)*	11	41.4 (3.7)	10	53.4 (4.5)	1.0[a]	0.000***	0.83
Number line task (% distance)	13	5.3 (1.9)	10	3.6 (2.2)	18.0[a]	0.002**	0.61
Addition (accuracy)	13	15.9 (4.0)	10	18.6 (1.4)	25.5[a]	0.010*	0.52
Subtraction (accuracy)	12	12.5 (4.1)	10	17.6 (2.5)	16.0[a]	0.002**	0.62
Domain general cognitive abilities							
Estimated IQ *(WISC-IV)*	13	100.5 (6.9)	10	113.0 (5.7)	9.5[a]	0.000***	0.72
Working memory							
Visuo-spatial *(BST)*	13	5.8 (1.8)	10	6.8 (2.0)	46.5[a]	0.243	0.25
Verbal *(WISC-IV)*	13	98.9 (12.1)	10	107.5 (9.2)	36.5[a]	0.074	0.37
Attention							
Alertness *(TAP)*	13	47.3 (11.5)	10	46.5 (10.4)	57.5[a]	0.659	0.10
Go-Nogo *(TAP)*	12	56.8 (32.9)	10	63.5 (24.1)	54.0[a]	0.710	0.08
Reading							
Words *(SLRT-II)*	12	19.1 (19.3)	10	12.9 (11.7)	55.5[a]	0.783	0.06
Pseudowords *(SLRT-II)*	11	24.5 (18.9)	10	19.3 (13.1)	45.0[a]	0.498	0.15

ZAREKI-R, Neuropsychological Test Battery for Number Processing and Calculation in Children (PR); BASIS-MATH 4-8, Basic Diagnostic in Mathematics for Grades 4-8 (raw score); WISC, Wechsler Intelligence Scale for Children (IQ score); KFT 4-12+R, Cognitive Abilities Test (T score); BST, Block-Suppression-Test (raw score); TAP, Testbattery for Attentional Performance (PR); SLRT-II, Salzburg Reading and Orthography Test (PR). [a]*Mann–Whitney U Test,* [b]*Fisher's Exact Test.* *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

differences. Effect sizes are reported as Pearson's correlation coefficient r and are interpreted as small ($r = 0.10$), medium ($r = 0.30$) or large ($r = 0.50$).

Brain Imaging

Image Acquisition

MRI data were acquired on a 3T General Electric Signa Scanner (GE Medical Systems, USA) using an 8-channel head coil. T1-weighted structural images (voxel size = $0.94 \times 0.94 \times 1.00$ mm^3) were acquired with a fast spoiled gradient echo sequence (3D FSPGR, slice thickness = 1 mm, no interslice skip, matrix size = 256×256, field of view = 240 mm, flip angle = 20°, echo time = 3 ms, repetition time = 10 ms). Participants were carefully instructed and supplied with hearing protection before entering the scanner. To minimize head motion, the head was stabilized with padding.

MRI Data Preprocessing

The data were preprocessed using the Computational Anatomy Toolbox (CAT12, Structural Brain Mapping Group, University of Jena, Germany), which is an extension to Statistical Parametric Mapping (SPM 12, Wellcome Trust Centre for Neuroimaging, University College London, UK) running on Matlab (Release 2012b, The MathWorks Inc., USA).

In a first step, the longitudinal data pair (baseline and follow-up image) of each subject was registered to the mean image for each subject by an inverse-consistent realignment, which also includes a bias correction between the different time points. The mean image of each subject is then segmented and the spatial normalization parameters are estimated with the help of a Dartel Normalization. These spatial normalization parameters (Dartel deformations) are then applied to the gray and the white matter segmentations of the baseline and follow-up image. The white and gray matter segmented images were smoothed with a Gaussian kernel of 8 mm FWHM (full width half maximum). As the last step, the total intracranial volume (TIV) was estimated for the baseline and follow-up image of each subject (please see **Supplementary Material** for the Matlab scripts of the preprocessing steps).

Quality Control

The CAT12 Toolbox provides image quality measures describing the properties of the images before the preprocessing. The image quality rating (IQR) is a weighted average of the noise contrast ratio, the inhomogeneity contrast ratio and the resolution of the input image. The images all reached ratings above 79 at both measurement time points (baseline range 79–87, follow-up range 81–87). Note that typical scientific (clinical) data is expected to get good to satisfactory ratings (70–90; Gaser and Dahnke, 2016, see also www.neuro.uni-jena.de/cat/). The groups did not differ in their IQR at baseline (DD *Mdn* = 85.0, TD *Mdn* = 85.3, $U = 46.5$, $z = -1.15$, $p = 0.263$) or follow-up (DD *Mdn* = 85.0, TD *Mdn* = 85.5, $U = 48.0$, $z = -1.05$, $p = 0.313$), and there was no significant difference in the quality measures between the time points (baseline *Mdn* = 85.1, follow-up *Mdn* = 85.3, $z = -0.69$, $p = 0.494$).

Also, the segmented and normalized gray and white matter images were visually inspected and the sample was checked for homogeneity (mean correlation). Based on the visual inspection and the Mahalanobis distance, which combines a measure of image quality before (weighted overall image quality) and after preprocessing (mean correlation), we excluded two DD and one TD data sets resulting in the final group size of 13 DD and 10 TD.

Statistical Model

For the statistical analyses of the gray and white matter volumes, two separate flexible factorial models with the factors subject, group (DD, TD) and time (baseline, follow-up) including TIV and puberty score as covariates were defined. Statistical results are shown with a threshold of $p < 0.05$ family-wise error (FWE) correction (see **Supplementary Material** for the design matrix and the Matlab scripts of the statistical model as well as the defined contrasts). Anatomical localization of the gray matter volume results was attained trough the SPM Anatomy Toolbox v2.0 (Eickhoff et al., 2005, 2007). White matter regions were labeled according to the JHU (Johns Hopkins University) white-matter tractography atlas (Hua et al., 2008).

RESULTS

Behavioral Data

The neuropsychological results and the demographic data for all subjects are summarized in **Table 1**. All participants scored in the normal range of IQ (DD IQ = 93–111, TD IQ = 101–125). However, groups differed in the estimated general IQ at baseline (WISC-III $p < 0.001$, $r = 0.69$) and follow-up (WISC-IV $p < 0.001$, $r = 0.72$). IQ measures are known not to be fully independent of measures of math ability, and the present sample, therefore, reflects the cognitive pattern typically observed in DD.

In the attention, working memory, and reading task no differences between DD and TD children were found (all $p > 0.05$).

Numerical Achievement

As expected, numerical abilities, assessed by the Zareki-R at baseline, differed significantly between the TD and the DD groups ($p < 0.001$, $r = 0.84$). The groups also differed in the subtest Arithmetic (WISC-II), with the DD subjects scoring significantly lower than the TD group ($p < 0.01$, $r = 0.55$; **Table 1**).

At follow-up, adolescents of the DD group still performed worse in comparison to their peers (Basis-Math $p < 0.001$, $r = 0.84$). In fact, all the subjects identified with DD at baseline still fulfilled the diagnostic criteria for DD at the follow-up measurement. Furthermore, they also scored significantly lower in the curriculum based test Quantity Comparison (KFT 4-12+R $p < 0.001$, $r = 0.83$), the number line task ($p < 0.01$, $r = 0.61$) and the basic arithmetic operations (addition $p < 0.05$, $r = 0.52$, subtraction $p < 0.01$, $r = 0.62$; **Table 1**).

Pearson's correlations of the whole brain gray and white matter volume with behavioral measures were calculated for baseline and follow-up. At the follow-up, gray matter volume was positively correlated with the Basis-Math ($r = 0.64$, $p < 0.05$) in TD children. However, we found a negative relationship in TD adolescents between performance in addition and gray ($r = -0.69$, both $p < 0.05$) and white matter volume ($r = -0.60$, $p < 0.05$), respectively. For DD children, no significant correlations between numerical abilities and volume of the brain structure were revealed (see **Supplementary Material** for a complete table of correlations).

Structural Results

Gray Matter

For the gray matter volume, the flexible factorial analysis revealed a significant effect of group (**Figure 1A**, **Table 2**). DD children showed decreased gray matter volumes in the bilateral inferior parietal lobe assigned to the IPS, the bilateral supramarginal gyri, the left precuneus, the left postcentral gyrus, and the right paracentral lobule compared to TD children. In the occipital lobe, differences were found in the left calcarine gyrus/cuneus, the left middle occipital gyrus (MOG), and the right superior occipital gyrus (SOG). Decreased gray matter was found in the bilateral inferior (ITG) and middle temporal gyri (MTG), the left rolandic operculum, and the bilateral insula over the examined time of 4 years.

The main effect of time and the interaction group by time were not significant.

White Matter

For the white matter volume, the flexible factorial analysis revealed a significant effect of the group (**Figure 1B**, **Table 3**). DD children showed reduced white matter volumes in a widespread set of brain regions including the bilateral CST, the bilateral superior and ILF, the bilateral IFOF, and the right ATR. Compared to TD children the DD children further showed increased white matter volumes in the right CST.

The main effect of time revealed an increase in the left SLF adjacent to the precentral gyrus over both groups. The interaction group by time was not significant.

DISCUSSION

Our study aimed to investigate the neural structural development of children with DD and TD peers using a longitudinal study. Until now, there have only been a handful of studies investigating the structural differences between children with and without DD, and only one study examining changes of regional differences

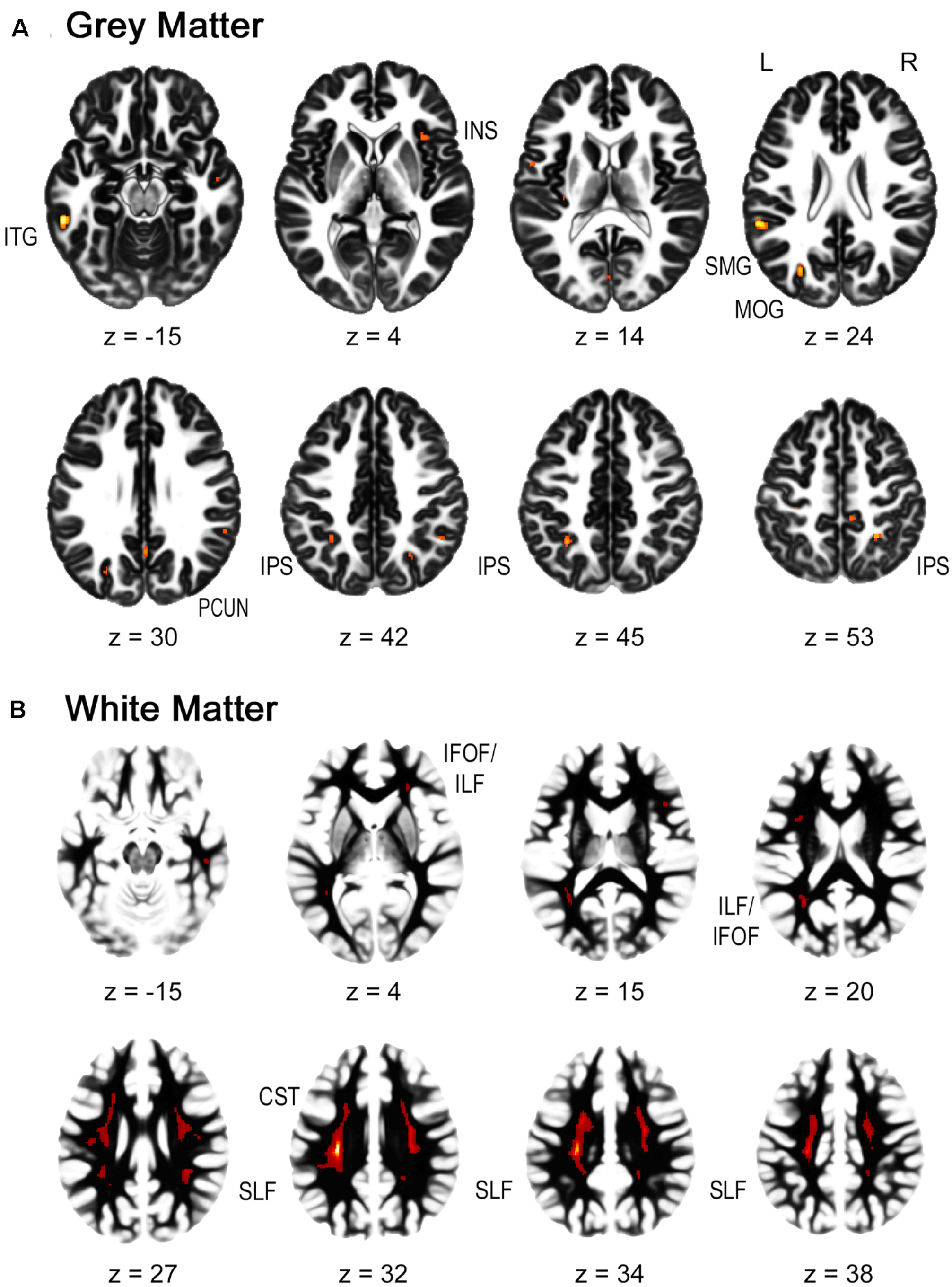


FIGURE 1 | Group differences in gray and white matter. Results are shown on a pediatric template (Fonov et al., 2009) with a significance level of p-cluster < 0.05, FWE corrected. **(A)** Decreased gray matter volumes in dyscalculics compared to typically developing children over both time points. **(B)** Decreased white matter volumes in dyscalculics compared to typically developing children over both time points. Abbreviations: CST, corticospinal tract; IFOF, inferior fronto-occipital fasciculus; ILF, inferior longitudinal fasciculus; INS, insula; IPS, intraparietal sulcus; ITG, inferior temporal gyrus; MOG, middle occipital gyrus; PCUN, precuneus; SLF, superior longitudinal fasciculus; SMG, supramarginal gyrus.

in cortical development (Rotzer et al., 2008; Rykhlevskaia et al., 2009; Ranpura et al., 2013). However, all of these study results are based on cross-sectional data. To our knowledge, this is the first study investigating the neural developmental trajectory using longitudinal data in children with DD. On the behavioral level, we found that the children of the DD group performed significantly worse in all the numerical and arithmetical tasks. This result remained stable over time. All children that were identified with DD at the beginning of the study still fulfilled the diagnostic criteria of DD 4 years later.

TABLE 2 | Gray matter.

Region	Cluster size	*p*-corrected	MNI coordinates			
			Z-value	*x*	*y*	*z*
L middle occipital gyrus	106	<0.001	6.04	−28	−75	27
L inferior temporal gyrus	101	<0.001	6.52	−58	−39	−15
L supramarginal gyrus/inferior parietal lobe	96	<0.001	6.18	−58	−40	24
L intraparietal sulcus/inferior parietal lobe	39	0.002	5.61	−32	−52	45
L precuneus	35	0.006	5.41	2	−58	30
R insula	31	0.008	5.35	36	20	4
L calcarine gyrus, cuneus	30	0.022	5.12	0	−82	9
L middle temporal gyrus/inferior parietal lobe	25	0.002	5.61	−56	−62	8
R paracentral lobe/superior parietal lobe	24	0.008	5.35	9	−36	51
R superior parietal lobe/intraparietal sulcus	23	<0.001	6.02	27	−51	54
R intraparietal sulcus/inferior parietal lobe	18	0.013	5.25	48	−50	40
R supramarginal gyrus/inferior parietal lobe	17	0.007	5.39	58	−45	28
L rolandic operculum	17	0.012	5.26	−56	2	14
R superior occipital gyrus	11	0.019	5.15	26	−63	44
R inferior temporal gyrus	9	0.024	5.10	44	−52	−10
R middle temporal gyrus	7	0.020	5.14	52	−9	−15
L postcentral gyrus	6	0.008	5.34	−32	−32	54
R insula	5	0.018	5.16	36	8	8
R middle temporal gyrus	4	0.039	4.98	42	−63	0
L insula	3	0.035	5.01	−33	−24	15
R insula	2	0.032	5.03	40	8	−3
L supramarginal gyrus/inferior parietal lobe	1	0.044	4.95	−58	−22	21

Peak coordinates and details of clusters from the whole brain voxel-based analyses (p < 0.05, FWE corrected).

TABLE 3 | White matter.

Region	Cluster size	*p*-corrected	MNI coordinates			
			Z-value	*x*	*y*	*z*
L corticospinal tract	1661	<0.001	>8.00	−28	−21	32
L superior longitudinal fasciculus		<0.001	>8.00	−21	−2	36
L superior longitudinal fasciculus		<0.001	7.46	−32	−9	28
R superior longitudinal fasciculus	677	<0.001	>8.00	27	−8	34
R superior longitudinal fasciculus		<0.001	>8.00	30	−21	32
N/A		<0.001	7.41	22	2	36
R superior longitudinal fasciculus	109	<0.001	7.39	34	−40	27
R anterior thalamic radiation	37	<0.001	6.54	22	−44	33
R superior longitudinal fasciculus	27	0.001	5.70	50	−27	−15
R corticospinal tract	25	0.008	5.23	16	−15	56
L inferior fronto-occipital fasciculus	22	0.005	5.34	−24	−84	−6
R inferior fronto-occipital fasciculus	17	0.009	5.20	26	33	4
R inferior longitudinal fasciculus/inferior fronto-occipital fasciculus	15	0.001	5.74	33	−68	9
N/A	9	0.002	5.48	44	20	15
L superior longitudinal fasciculus	2	0.039	4.84	−20	−46	51
L inferior longitudinal fasciculus	2	0.015	5.08	−46	−14	−21
L superior longitudinal fasciculus	1	0.046	4.80	−44	−28	28
N/A	1	0.045	4.81	34	−78	14
L inferior longitudinal fasciculus	1	0.043	4.82	−32	−80	−4
R inferior longitudinal fasciculus	1	0.040	4.84	46	−40	−8

Peak coordinates and details of clusters from the whole brain voxel-based analyses (p < 0.05, FWE corrected).

On the neural level, children with DD showed reduced gray and white matter volumes in various regions and prominent tracts of the frontoparietal numerical network. These differences do not vanish over time, but persist from childhood into adolescence.

The dyscalculics showed reduced gray matter volumes in the parietal lobes specifically, but also in the occipital, temporal, and frontal parts of the brain, consistent with the results of previous studies. Less gray matter volume in the IPL including the IPS and the SPL have been reported in all studies with dyscalculic children (Rotzer et al., 2008; Rykhlevskaia et al., 2009; Ranpura et al., 2013). These regions are known from functional studies to be the key areas for number processing and quantity representation (for a meta-analysis see Sokolowski et al., 2017; Arsalidou et al., 2018). Furthermore, we also found reduced gray matter volumes in the bilateral supramarginal gyri, which are thought to play a crucial role in the retrieval of arithmetical facts (Menon, 2016). Similar to Rykhlevskaia et al. (2009), our results revealed lower gray matter volumes in the MOG/SOG, the cuneus/precuneus, and the temporal gyrus, although our results include bilateral MTG/ITG. Unlike the studies conducted before with DD children (Rotzer et al., 2008; Rykhlevskaia et al., 2009; Ranpura et al., 2013), we found reduced gray matter volumes in the insula. However, the insula has a high likelihood to be activated when children solve number and calculation tasks and have been proposed to play a role in intrinsic motivation about learning and training (Arsalidou et al., 2018). Also, our study did not find any volumetric differences in the parahippocampal areas, which was reported in Rykhlevskaia et al. (2009).

In terms of white matter, our study revealed reduced white matter volume in dyscalculics in the bilateral superior and ILF, the CST, the IFOF, and the ATR. Reduced white matter volumes were also reported in the same tracts by Rykhlevskaia et al. (2009), except that they found additionally reduced volumes in the forceps major and the splenium of the corpus callosum. Moreover, the CST, the ILF and SLF, and the corona radiata (of which the ATR is part) have all been associated with numerical and mathematical processing by numerous studies (Kucian et al., 2013; for a review see also Matejko and Ansari, 2015). The SLF connects frontal and parietal regions of the brain, which are known to be the main areas activated when solving number and arithmetic related tasks. It has further been proposed that the ILF is involved in visual processing related to numerical or mathematical problem solving (van Eimeren et al., 2008; Matejko and Ansari, 2015).

Children with DD revealed increased volumes in the right CST. This tract connects the cortex with the brainstem and is typically associated with motor functions. Research has further demonstrated that there is a link between finger and number representation (Noël, 2005; Matejko and Ansari, 2015). As children with DD often rely on finger counting strategies to compensate for the deficits in fact retrieval, the increased volume in the CST might be related to the frequent finger use during calculation in DD.

However, it is important to note that these observed structural differences in gray and white matter are related to underlying microstructural mechanisms of development and learning. Among the candidate mechanisms explaining gray and white matter plasticity are morphometric changes in the neuron (e.g., axon sprouting, dendritic branching, synaptogenesis, neurogenesis), in fiber organization (e.g., axon branching, sprouting, axon diameter or the number of axons) as well as in the myelination (for detailed information see Zatorre et al., 2012). But also vascular changes (angiogenesis) or changes in morphology and number of glia and astrocytes could explain the increase in gray and white matter volume (Zatorre et al., 2012). Regarding the structural differences in DD children, one could speculate that DD children show reduced white and gray matter volumes as the underlying microstructural process does not take place to the same extent as in TD peers. This would also be in line with results of functional and DTI studies reporting decreased activation and lower FA values in number related areas and tracks, respectively (Davis et al., 2009; Kucian et al., 2013; for an overview see Peters and De Smedt, 2018).

Over development, we did not observe a gray matter decrease or prominent white matter increases. The lack of a gray matter decrease can be due to the age of our subjects. We examined children between the ages of 9 and 14. The developmental trajectory of the gray matter volume follows an inverted u-shape and depending on the study the peak of gray matter volume has been reported at age 8 (Mills et al., 2016), age 12 (Groeschel et al., 2010), or between ages 11–14 (Gogtay and Thompson, 2010). The time point of the peak further varies depending on the brain region (Giedd et al., 1999) and the sex and/or pubertal stage, with females reaching gray matter peaks 1–2 years earlier than males (Gogtay and Thompson, 2010; Mills and Tamnes, 2014). Therefore, it could be the case that we did not detect developmental changes in the gray matter, because the children and adolescents we studied are around one of the reported gray matter peaks. If we look at the individual trajectories of our study participants, some children still show a gray matter increase, whilst other children show a gray matter decrease or almost no change in the gray matter volume over the 4 years (**Figure 2**, upper panel). Moreover, the fact that we controlled for pubertal status and therefore indirectly for sex might be an additional reason why we do not find developmental changes. In the white matter volume, we found a significant increase during development in the left SLF (MNI $x = -40$, $y = -20$, $z = 28$), located right next to the reported developmental changes in the study of Giorgio et al. (2010; MNI $x = -42$, $y = -22$, $z = 28$). However, our results do not show the prominent white matter changes as reported in the literature (Giorgio et al., 2010). A possible explanation is that our study does not look at the correlations between white matter volume and age. On the other hand, the individual trajectories of our subjects show a clear increase in the total white matter volume (**Figure 2**, lower panel), which is in line with previous research (Aubert-Broche et al., 2013; Mills et al., 2016).

The main focus of our study was to find out more about the developmental trajectories of gray and white matter volumes in children with DD. We found no significant interactions, but the group differences in gray and white matter volumes remained stable during the examined time window. Although the differences found between DD and TD children are well in line with the literature (Rotzer et al., 2008; Rykhlevskaia et al., 2009), and also reflect the known and persistent behavioral differences (Nelson and Powell, 2018), the developmental trajectories differ from the results reported by Ranpura et al. (2013) in such that we did not find an increase in white matter volume in frontal and parietal areas in TD children only. Studies in adults find no or very subtle differences between the subjects with and without DD (Cappelletti and Price, 2013; Moreau et al., 2019). Based on these results, one could argue that the volumetric differences should diminish over development. However, Moreau et al. (2019) used a rather lenient criterion for DD, which could also be a reason why volumetric differences were not found. Also, there is an age gap of 10–15 years between our results and the adult studies, in which developmental changes can still take place. Our results, therefore, point towards stable and persistent differences in the

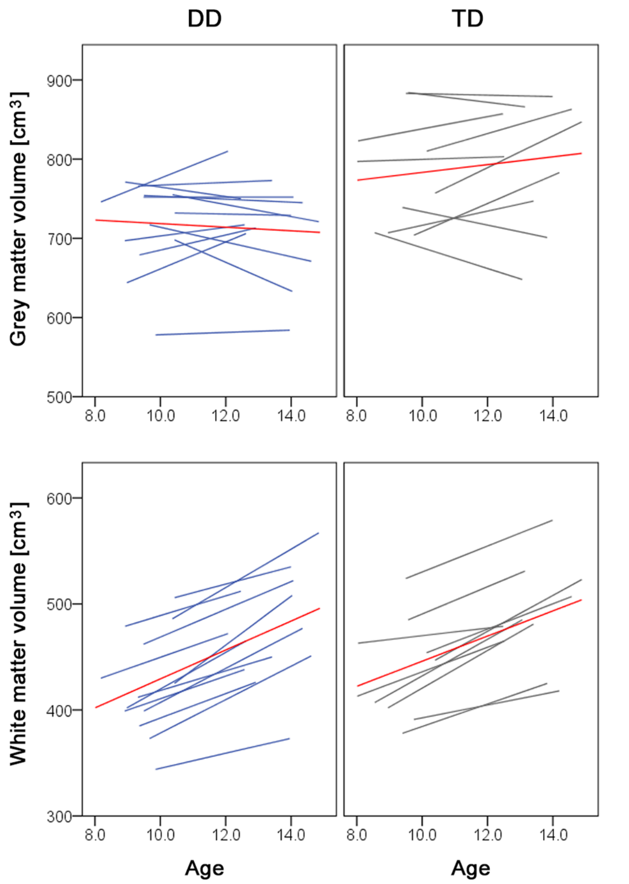


FIGURE 2 | Spaghetti plots of whole brain volumes. Gray matter (top panel) and white matter volumes (lower panel) are plotted against age for children with dyscalculia (left, in blue) and typically developing children (right, in gray). The red line depicts the mean of the group.

dyscalculics gray and white matter volumes from childhood to adolescence. More research, specifically longitudinal studies over a longer time interval, are urgently needed to enable us to conclude the developmental trajectory of the brain structure in DD. However, findings from the present study suggest that development proceeds in a similar manner between DD and TD children.

To better understanding the present findings, it is important that we advance our knowledge about the typical developmental trajectory of the brain structure and its spontaneous variations in numerical cognition. Furthermore, the effects of schooling and specific interventions on brain structure need to be explored more profoundly. This knowledge should build the basis for the investigation and a better understanding of the deviant and/or delayed development as reported in children with DD. In a next step, it would be interesting to investigate if the abnormalities in gray and white matter are a result of a developmental delay or a specific marker of DD. This open question could be tackled by comparing the structural brain development of DD children with TD children that perform on a similar numerical level. Furthermore, the investigation of structural changes caused by a specific intervention could help clarify the questions about the neurobiological cause of dyscalculia.

Limitations

Our results are important for the field since this is to our knowledge the first study investigating the developmental trajectory of structural white and gray matter volumes using longitudinal data in DD. However, several limitations should be considered when interpreting the results of our study. First of all, due to braces and movement artifacts, the drop-out rates in longitudinal MRI studies with children and adolescents are high. Our study includes only small sample sizes and should for this reason be interpreted with caution. However, the data included had good data quality ratings, as assessed by objective criteria (Gaser and Dahnke, 2016). Moreover, we replicated the main findings of previous studies examining volumetric gray and white matter differences in DD children (Rotzer et al., 2008; Rykhlevskaia et al., 2009; Ranpura et al., 2013), which were also performed with larger sample sizes (e.g., Rykhlevskaia et al., 2009). Therefore, we are confident that despite the small sample sizes our results contribute valuable knowledge towards an understanding of the developmental trajectory of brain structure in children with and without DD.

Second, the rather large age range within our sample (at each time point) may attenuate the developmental effects. For methodological reasons, a narrower age range would be much better to examine general developmental effects and detect group differences over time. A closer look at our data revealed that the individual developmental trajectories showed similar trends irrespective of the age at the entry of the study. However, future longitudinal studies should investigate limited age ranges to control better for the effect of schooling and the rapid changes in development.

A third restricting point of our study is the significant difference in IQ between the TD and the DD group. Developmental imaging studies have shown that there is a positive relationship between intellectual abilities and white/gray matter volume, especially in the dorsolateral prefrontal cortex, parietal lobe, the anterior cingulate cortex and in temporal and occipital regions (Wilke et al., 2003; Tamnes et al., 2011; Brancucci, 2012). For this reason, we ran our analyses with IQ as an additional covariate showing that the main results remained unchanged (see **Supplementary Material**).

Last, it should be noted that some authors argue that the sensitivity to examine the white matter using voxel-based morphometry is limited, as white matter areas are characterized by large homogenous regions with only subtle changes in intensity (Kurth et al., 2015).

Conclusion

In conclusion, the present study reveals for the first time the gray and white matter trajectories of the dyscalculic brain. The findings confirm the structural differences as reported in earlier research and support the notion that DD is characterized by persistent structural and behavioral abnormalities. There is an urgent need for longitudinal studies examining the typical and atypical neural development, but also the effect of interventions and therapy on numerical and mathematical abilities. Advancing the knowledge about the developmental course of DD and the effects of schooling, therapy, and intervention would enable us to support affected children and adolescents more effectively.

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available. According to the Ethics committee of Zurich, Switzerland the consent was not obtained for sharing data outside the research team. Therefore we do not have the allowance to make the data publicly available. Requests to access the datasets should be directed to the Ethics committee of Zurich in Switzerland (Info.KEK@kek.zh.ch, project-number: 2011-0384, title of the project: Plasticity after training and development in children with and without dyscalculia).

ETHICS STATEMENT

The present study was reviewed and approved by the Ethics committee of Canton of Zurich, Switzerland. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

All authors have contributed and have approved the final manuscript. UM contributed to the design of the study, the acquisition, analysis, interpretation of the data, and writing the manuscript. MA and RO'G contributed to data interpretation and revised the manuscript. KK contributed to the design of the study, the acquisition, data interpretation, editing, and revision of the manuscript.

FUNDING

This research was supported by a grant from the NOMIS Foundation.

ACKNOWLEDGMENTS

We would like to thank all subjects and their parents for participating in this study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2020.00272/full#supplementary-material.

REFERENCES

Arsalidou, M., Pawliw-Levac, M., Sadeghi, M., and Pascual-Leone, J. (2018). Brain areas associated with numbers and calculations in children: meta-analyses of fMRI studies. *Dev. Cogn. Neurosci.* 30, 239–250. doi: 10.1016/j.dcn.2017.08.002

Ashkenazi, S., Rosenberg-Lee, M., Tenison, C., and Menon, V. (2012). Weak task-related modulation and stimulus representations during arithmetic problem solving in children with developmental dyscalculia. *Dev. Cogn. Neurosci.* 2, S152–S166. doi: 10.1016/j.dcn.2011.09.006

Aubert-Broche, B., Fonov, V. S., García-Lorenzo, D., Mouiha, A., Guizard, N., Coupé, P., et al. (2013). A new method for structural volume analysis of longitudinal brain MRI data and its application in studying the growth trajectories of anatomical brain structures in childhood. *NeuroImage* 82, 393–402. doi: 10.1016/j.neuroimage.2013.05.065

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

Brancucci, A. (2012). Neural correlates of cognitive ability. *J. Neurosci. Res.* 90, 1299–1309. doi: 10.1002/jnr.23045

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

Cappelletti, M., and Price, C. J. (2013). Residual number processing in dyscalculia. *NeuroImage Clin.* 4, 18–28. doi: 10.1016/j.nicl.2013.10.004

Carskadon, M. A., and Acebo, C. (1993). A self-administered rating scale for pubertal development. *J. Adolesc. Health* 14, 190–195. doi: 10.1016/1054-139x(93)90004-9

Davis, N., Cannistraci, C. J., Rogers, B. P., Gatenby, J. C., Fuchs, L. S., Anderson, A. W., et al. (2009). Aberrant functional activation in school age children at-risk for mathematical disability: a functional imaging study of simple arithmetic skill. *Neuropsychologia* 47, 2470–2479. doi: 10.1016/j.neuropsychologia.2009.04.024

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 skills 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

Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M.-H., Evans, A. C., Zilles, K., et al. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *NeuroImage* 36, 511–521. doi: 10.1016/j.neuroimage.2007.03.060

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

Evans, T. M., Kochalka, J., Ngoon, T. J., Wu, S. S., Qin, S., Battista, C., et al. (2015). Brain structural integrity and intrinsic functional connectivity forecast 6 year longitudinal growth in children's numerical abilities. *J. Neurosci.* 35, 11743–11750. doi: 10.1523/jneurosci.0216-15.2015

Fonov, V., Evans, A., McKinstry, R., Almli, R., and Collins, L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage* 47, S39–S41. doi: 10.1016/s1053-8119(09)70884-5

Gaser, C., and Dahnke, R. (2016). CAT-A computational anatomy toolbox for the analysis of structural MRI data. *OHBM* 2016, 336–348.

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., 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

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., et al. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nat. Neurosci.* 2, 861–863. doi: 10.1038/13158

Giorgio, A., Watkins, K. E., Chadwick, M., James, S., Winmill, L., Douaud, G., et al. (2010). Longitudinal changes in grey and white matter during adolescence. *NeuroImage* 49, 94–103. doi: 10.1016/j.neuroimage.2009.08.003

Gogtay, N., and Thompson, P. M. (2010). Mapping gray matter development: implications for typical development and vulnerability to psychopathology. *Brain Cogn.* 72, 6–15. doi: 10.1016/j.bandc.2009.08.009

Groeschel, S., Vollmer, B., King, M. D., and Connelly, A. (2010). Developmental changes in cerebral grey and white matter volume from infancy to adulthood. *Int. J. Dev. Neurosci.* 28, 481–489. doi: 10.1016/j.ijdevneu.2010.06.004

Heller, K. A., and Perleth, C. (2000). *Cognitive Abilities Test for Grades 4 to 12 - Revised Version.* Göttingen: Beltz.

Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D. S., et al. (2008). Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *NeuroImage* 39, 336–347. doi: 10.1016/j.neuroimage.2007.07.053

Isaacs, E. B., Edmonds, C. J., Lucas, A., and Gadian, D. G. (2001). Calculation difficulties in children of very low birthweight: a neural correlate. *Brain* 124, 1701–1707. doi: 10.1093/brain/124.9.1701

Kronschnabel, J., Schmid, R., Maurer, U., and Brandeis, D. (2013). Visual print tuning deficits in dyslexic adolescents under minimized phonological demands. *NeuroImage* 74, 58–69. doi: 10.1016/j.neuroimage.2013.02.014

Kucian, K., Ashkenazi Schwizer, S., Hänggi, J., Rotzer, S., Jäncke, L., Martin, E., et al. (2013). Developmental dyscalculia: a dysconnection syndrome? *Brain Struct. Funct.* 219, 1721–1733. doi: 10.1007/s00429-013-0597-4

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

Kurth, F., Luders, E., and Gaser, C. (2015). "Voxel-based morphometry," in *Brain Mapping*, ed. A. W. Toga (Waltham, MA: Academic Press), 345–349.

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

Li, Y., Hu, Y., Wang, Y., Weng, J., and Chen, F. (2013). Individual structural differences in left inferior parietal area are associated with schoolchildrens' arithmetic scores. *Front. Hum. Neurosci.* 7:844. doi: 10.3389/fnhum.2013.00844

Matejko, A. A., and Ansari, D. (2015). Drawing connections between white matter and numerical and mathematical cognition: a literature review. *Neurosci. Biobehav. Rev.* 48, 35–52. doi: 10.1016/j.neubiorev.2014.11.006

McCaskey, U., von Aster, M., Maurer, U., Martin, E., O'Gorman Tuura, R., and Kucian, K. (2018). Longitudinal brain development of numerical skills in typically developing children and children with developmental dyscalculia. *Front. Hum. Neurosci.* 11:629. doi: 10.3389/fnhum.2017.00629

Menon, V. (2016). Working memory in children's math learning and its disruption in dyscalculia. *Curr. Opin. Behav. Sci.* 10, 125–132. doi: 10.1016/j.cobeha.2016.05.014

Mills, K. L., Goddings, A.-L., Herting, M. M., Meuwese, R., Blakemore, S.-J., Crone, E. A., et al. (2016). Structural brain development between childhood and adulthood: convergence across four longitudinal samples. *NeuroImage* 141, 273–281. doi: 10.1016/j.neuroimage.2016.07.044

Mills, K. L., and Tamnes, C. K. (2014). Methods and considerations for longitudinal structural brain imaging analysis across development. *Dev. Cogn. Neurosci.* 9, 172–190. doi: 10.1016/j.dcn.2014.04.004

Moll, K., and Landerl, K. (2010). *Salzburg Reading and Orthography Test - Revised version.* Bern: Huber.

Moreau, D., Wiebels, K., Wilson, A. J., and Waldie, K. E. (2019). Volumetric and surface characteristics of gray matter in adult dyslexia and dyscalculia. *Neuropsychologia* 127, 204–210. doi: 10.1016/j.neuropsychologia.2019.02.002

Moser Opitz, E., Reusser, L., Moeri Müller, M., Anliker, B., Wittich, C., and Freesemann, O. (2010). *BASIS-MATH 4–8. Basisdiagnostik Mathematik für die Klassen 4-8 [Basic Diagnosis in Mathematics Education for Grades 4-8].* Bern: Huber.

Nelson, G., and Powell, S. R. (2018). A systematic review of longitudinal studies of mathematics difficulty. *J. Learn. Disabil.* 51, 523–539. doi: 10.1177/0022219417714773

Noël, M.-P. (2005). Finger gnosia: a predictor of numerical abilities in children? *Child Neuropsychol.* 11, 413–430. doi: 10.1080/09297040590951550

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Petermann, F., and Petermann, U. (2007). *Wechsler Intelligence Scale for Children - Fourth Edition.* Bern: Huber.

Peters, L., and De Smedt, B. (2018). Arithmetic in the developing brain: a review of brain imaging studies. *Dev. Cogn. Neurosci.* 30, 265–279. doi: 10.1016/j.dcn.2017.05.002

Popescu, T., Sader, E., Schaer, M., Thomas, A., Terhune, D. B., Dowker, A., et al. (2019). The brain-structural correlates of mathematical expertise. *Cortex* 114, 140–150. doi: 10.1016/j.cortex.2018.10.009

Price, G. R., Holloway, I., Räsänen, P., Vesterinen, M., and Ansari, D. (2007). Impaired parietal magnitude processing in developmental dyscalculia. *Curr. Biol.* 17, R1042–R1043. doi: 10.1016/j.cub.2007.10.013

Price, G. R., Wilkey, E. D., Yeo, D. J., and Cutting, L. E. (2016). The relation between 1st grade grey matter volume and 2nd grade math competence. *NeuroImage* 124, 232–237. doi: 10.1016/j.neuroimage.2015.08.046

Ranpura, A., Isaacs, E., Edmonds, C., Rogers, M., Lanigan, J., Singhal, A., et al. (2013). Developmental trajectories of grey and white matter in dyscalculia. *Trends Neurosci. Educ.* 2, 56–64. doi: 10.1016/j.tine.2013.06.007

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

Rykhlevskaia, E., Uddin, L. Q., Kondos, L., and Menon, V. (2009). Neuroanatomical correlates of developmental dyscalculia: combined evidence from morphometry and tractography. *Front. Hum. Neurosci.* 3:51. doi: 10.3389/neuro.09.051.2009

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. doi: 10.1017/s0012162205000216

Sokolowski, H. M., Fias, W., Mousa, A., and Ansari, D. (2017). Common and distinct brain regions in both parietal and frontal cortex support symbolic and nonsymbolic number processing in humans: a functional neuroimaging meta-analysis. *NeuroImage* 146, 376–394. doi: 10.1016/j.neuroimage.2016.10.028

Starke, M., Kiechl-Kohlendorfer, U., Kucian, K., Pupp Peglow, U., Kremser, C., Schocke, M., et al. (2013). Brain structure, number magnitude processing and math proficiency in 6- to 7-year-old children born prematurely: a voxel-based morphometry study. *Neuroreport* 24, 419–424. doi: 10.1097/wnr.0b013e32836140ed

Supekar, K., Swigart, A. G., Tenison, C., Jolles, D. D., Rosenberg-Lee, M., Fuchs, L., et al. (2013). Neural predictors of individual differences in response to math tutoring in primary-grade school children. *Proc. Natl. Acad. Sci. U S A* 110, 8230–8235. doi: 10.1073/pnas.1222154110

Szucs, D., Devine, A., Soltesz, F., Nobes, A., and Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex* 49, 2674–2688. doi: 10.1016/j.cortex.2013.06.007

Tamnes, C. K., Fjell, A. M., Østby, Y., Westlye, L. T., Due-Tønnessen, P., Bjørnerud, A., et al. (2011). The brain dynamics of intellectual development: waxing and waning white and gray matter. *Neuropsychologia* 49, 3605–3611. doi: 10.1016/j.neuropsychologia.2011.09.012

Tewes, U., Rossmann, P., and Schallberger, U. (1999). *Wechsler Intelligence Scale for Children - Third Edition.* Bern: Huber.

Tsang, J. M., Dougherty, R. F., Deutsch, G. K., Wandell, B. A., and Ben-Shachar, M. (2009). Frontoparietal white matter diffusion properties predict mental arithmetic skills in children. *Proc. Natl. Acad. Sci. U S A* 106, 22546–22551. doi: 10.1073/pnas.0906094106

van Eimeren, L., Niogi, S., McCandliss, B., Holloway, I., and Ansari, D. (2008). White matter microstructures underlying mathematical abilities in children. *Neuroreport* 19, 1117–1121. doi: 10.1097/wnr.0b013e328307f5c1

von Aster, M., Weinhold Zulauf, M., and Horn, R. (2006). *ZAREKI-R. Neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern, Revidierte Version [Neuropsychological Test Battery for Number Processing and Calculation in Children].* Frankfurt am Main: Harcourt.

WHO. (2018). International classification of diseases for mortality and morbidity statistics (ICD-11). Available online at: https://icd.who.int/browse11/l-m/en. Accessed November 4, 2019.

Wilke, M., Sohn, J.-H., Byars, A. W., and Holland, S. K. (2003). Bright spots: correlations of gray matter volume with IQ in a normal pediatric population. *NeuroImage* 20, 202–215. doi: 10.1016/s1053-8119(03)00199-x

Wilkey, E. D., Cutting, L. E., and Price, G. R. (2018). Neuroanatomical correlates of performance in a state-wide test of math achievement. *Dev. Sci.* 21:2. doi: 10.1111/desc.12545

WMA. (2013). *The World Medical Association's Declaration of Helsinki—Ethical Principles for Medical Research Involving Human Subjects.* Available online at: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/. Accessed December 20, 2017

Wyschkon, A., Kohn, J., Ballaschk, K., and Esser, G. (2009). Sind Rechenstörungen genau so häufig wie Lese-Rechtschreibstörungen? [Is a specific disorder of arithmetic skills as common as reading/spelling disorder?]. *Z. Für Kinder Jugendpsych. Psychother.* 37, 499–510. doi: 10.1024/1422-4917.37.6.499

Zatorre, R. J., Fields, R. D., and Johansen-Berg, H. (2012). Plasticity in gray and white: neuroimaging changes in brain structure during learning. *Nat. Neurosci.* 15, 528–536. doi: 10.1038/nn.3045

Zhao, Q., Zhang, Z., Xie, S., Pan, H., Zhang, J., Gong, G., et al. (2013). Cognitive impairment and gray/white matter volume abnormalities in pediatric patients with Turner syndrome presenting with various karyotypes. *J. Pediatr. Endocrinol. Metab.* 26, 1111–1121. doi: 10.1515/jpem-2013-0145

Zimmermann, P., and Fimm, B. (1993). *Testbattery for Attentional Performance.* Freiburg: Psytest.

Conflict of Interest: 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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 30 October 2020
doi: 10.3389/fpsyg.2020.489371

Self-Regulation and Mathematics Performance in German and Iranian Students of More and Less Math-Related Fields of Study

Parvin Nemati[1], Caterina Gawrilow[1,2], Hans-Christoph Nuerk[1,2,3] and Jan Kühnhausen[4]*

[1] Department of Psychology, University of Tübingen, Tübingen, Germany, [2] LEAD Graduate School & Research Network, University of Tübingen, Tübingen, Germany, [3] Leibniz-Institut für Wissensmedien, Tübingen, Germany, [4] Department for Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University of Tübingen, Tübingen, Germany

Edited by:
Reuven Babai,
Tel Aviv University, Israel

Reviewed by:
Anna Mulasso,
University of Turin, Italy
Susana Rodríguez,
University of A Coruña, Spain

***Correspondence:**
Parvin Nemati
parvin.nemati@uni-tuebingen.de;
pariyanemati@gmail.com
orcid.org/0000-0003-0098-4655

Specialty section:
This article was submitted to
Cognitive Science,
a section of the journal
Frontiers in Psychology

Received: 02 August 2019
Accepted: 05 October 2020
Published: 30 October 2020

Citation:
Nemati P, Gawrilow C, Nuerk H-C and Kühnhausen J (2020) Self-Regulation and Mathematics Performance in German and Iranian Students of More and Less Math-Related Fields of Study. Front. Psychol. 11:489371. doi: 10.3389/fpsyg.2020.489371

Self-regulation is a multidimensional construct that is positively related to academic achievement, such as successful mathematics performance. However, this relation of self-regulation and mathematics performance has mainly been investigated in Western countries with similar cultural contexts, although self-regulation is assumed to be context-sensitive. Therefore, the present study investigated the relation of self-regulation and mathematics performance across two different countries (Germany vs. Iran) in college students. The relation of self-regulation and mathematics performance was expected to be weaker in students of math-related fields, such as Engineering/Informatics, as they are assumed to need less self-regulation to solve the mathematics problems than students of less math-related fields, such as Human Sciences. In total, 122 undergraduate students (German = 60; Iranian = 62) of Human Sciences or Engineering/Informatics participated in this study. We measured self-regulation with the Brief Self-Control Scale (Tangney et al., 2004) and mathematics performance with a complex multiplication test. Results showed that self-regulation did not predict multiplication performance in German or Iranian students, in general. However, when the field of study was considered, self-regulation predicted multiplication performance in the subgroup of German and Iranian students studying Human Sciences within each country. We conclude that cultural context does not seem to play a dominant role in moderating the relation between self-regulation and math performance, however, field of study and more generally familiarity with math may be an important factor to consider in single or cross-cultural studies.

Keywords: self-regulation, mathematics, cross-culture, field of study, multiplication

INTRODUCTION

Self-regulation is defined as the ability to control one's thoughts, behaviors, or emotions, and enables individuals to adapt their behaviors in accordance with the demands of a situation (e.g., Baumeister and Vohs, 2007; Blair and Ursache, 2011). It includes abilities such as maintaining attention and inhibiting irrelevant information in learning situations, which provides an important foundation for successful academic outcomes (e.g., McClelland and Cameron, 2011). A large body of research connects self-regulation with different academic achievements, such as successful mathematics performance (e.g., Zimmerman, 1990; Bull and Scerif, 2001; Camahalan, 2006; Fuchs et al., 2006; Blair and Razza, 2007; Labuhn et al., 2010; McClelland et al., 2010; Otts, 2010; von Suchodoletz and Gunzenhauser, 2013; Gawrilow et al., 2014). For instance, college students with better self-regulation abilities measured by self-reports have been shown to respond more rapidly in mathematics tasks, which could be because of their enhanced ability to ignore distracting thoughts and concentrate on the task (Nemati et al., 2017). In contrast, students without adequate self-regulatory skills are more likely to experience difficulties in mathematics performance. For example, students who struggle with self-regulation, such as students with attention deficit/hyperactivity disorder (ADHD) have more difficulty with mathematics at school (e.g., Frazier et al., 2007; Zentall, 2007).

Previous studies have indicated that self-regulation contributes to mathematical performance by suppressing distracting thoughts or information whilst mathematics problems are solved (e.g., Gawrilow et al., 2011; Nemati et al., 2017), and through different cognitive components of self-regulation such as inhibitory control (e.g., Hofmann et al., 2011; McClelland and Cameron, 2011). For instance, solving complex multiplication problems requires ignoring distracting thoughts to remain focused on the task and selecting the correct solutions while suppressing alternative ones (e.g., neighboring solutions in the multiplication table) that can interfere with the retrieval of a desired solution (e.g., "42" can interfere with retrieving the answer to "6×8"; cf. Domahs et al., 2006, 2007).

However, the relation of self-regulation and mathematics performance might vary across different contexts. Recent studies demonstrated that self-regulation is a context-specific construct (e.g., Keller et al., 2004; von Suchodoletz et al., 2015; Lamm et al., 2018), suggesting that context can influence self-regulation displayed in different situations. For instance, the different parenting styles of European American and Puerto Rican mothers resulted in different patterns of self-regulation development during childhood (Carlson and Harwood, 2003): in the European American context, mothers expected their children to alter their behavior to match their individual goals, while Puerto Rican mothers asked their children to adjust their behavior in accordance to the society.

These findings are in line with the theoretical framework of Markus and Kitayama (1991), suggesting *independent and interdependent* contexts, which can influence self-regulation. Independent contexts focus on autonomy and individual goals, whereas interdependent contexts are associated with being in harmony with the group and the community goals. Accordingly, self-regulation processes in an independent context are directed toward influencing the environment and other people in line with an individual's goals, while in interdependent contexts they focus on adjusting one's behavior to the expectations of others to maintain fit with the group (Trommsdorff, 2009). For instance, the results of a recent cross-cultural study (Lamm et al., 2018) revealed that the development of self-regulation and self-regulatory strategies used by children can be different in independent and interdependent contexts. They showed that while German mothers emphasized autonomy and individual goals of their children, Cameroonian mothers expected their children to behave in harmony with society. Thus, German children's self-regulation was motivated by a different goal (i.e., autonomy in Germany vs. parents' expectations and group harmony in Cameroon) and for the same reason, German children might have used different self-regulatory strategies than their Cameroonian peers to do the self-regulation task.

Previous studies have showed that independent contexts are a core characteristic of Central European and North American countries, while interdependent contexts prevail in Asian and Latin American countries (e.g., Higgins et al., 2008; Trommsdorff, 2009). In the same line, individualism and autonomy are valued in Germany, while collectivism and group harmony are respected in Iran (Hofstede, 1980). Therefore, Germany and Iran provide two different contexts with distinct environmental characteristics that can affect self-regulation and its correlates.

However, although self-regulation has been frequently shown to have a context-sensitive nature (Trommsdorff, 2009; see also the review by Jaramillo et al., 2017), less is known about the relation of self-regulation and academic achievement, such as mathematics performance, across different countries and the existing results in children are rather scarce and heterogeneous. On the one hand, results of a cross-cultural study in preschool children demonstrated that the associations between different components of self-regulation and mathematics performance were largely similar between Chinese and North American children (Lan et al., 2011). They discussed that their finding might be due to the similarities in the associations between different cognitive components of self-regulation in distinct contexts. On the other hand, results of a longitudinal study investigating the application of self-regulatory strategies in educational settings, showed that many of the self-regulatory strategies used by Italian students did not predict the academic achievement as they did in American students (Nota et al., 2004). Researchers examined the self-regulatory strategies adopted by Italian students during the final year of high school and their academic achievement in pursuing further education at the University and compared their results with previous studies in American students. In the same vein, but in contrast to previous studies in Western countries, results of another study on Chinese students revealed no relationship between self-regulation and mathematics achievement in Chinese high school students (Rao et al., 2000). The authors suggested that self-regulatory strategies motivated by Chinese attitudes toward academic achievement and parents' expectations could not predict mathematics performance in Chinese high school students. Therefore, self-regulatory strategies adopted by students might not be equally important in predicting mathematics achievements across different countries. Altogether, it seems that independent and interdependent contexts can potentially impact the relation of self-regulation and mathematics performance. Furthermore, differences in self-regulatory skills across different countries can persist in adolescence (e.g., Ellefson et al., 2017), suggesting that context may influence self-regulation and its subsequent relationship with future academic, in particular, mathematics performance. Therefore, the aim of the present study was to examine whether the relation of self-regulation and mathematics performance varies between German and Iranian college students.

Additionally, field of study was considered as another context beside the country that could influence the relationship between self-regulation and mathematics performance in college students. It has been shown that individuals need more self-regulation when doing difficult tasks (e.g., Kanfer and Ackerman, 1989; Steele-Johnson et al., 2000) and solving mathematics problems might be less difficult for students of math-related fields, as compared to students of less math-related fields. Accordingly, context of field of study might affect the relationship between self-regulation and mathematics performance: the relationship between self-regulation and mathematics performance was expected to be weaker in students of math-related fields, such as Engineering/Informatics, as they are assumed to need less self-regulation to solve the mathematics problems than students of less math-related fields, such as Human Sciences. Therefore, the context-effect of field of study was taken into account in the present study as it can influence the students' mathematics performance and hence alter its relationship with self-regulation.

To sum up, in the present study, we hypothesized that the relation of self-regulation and mathematics performance differs in German and Iranian college students as independent and interdependent contexts can differentially affect self-regulation and its correlates. Furthermore, as the second hypothesis, we expected that the relation of self-regulation and mathematics performance is weaker in students of math-related fields, such as Engineering/Informatics, than in students of less math-related fields, such as Human Sciences, because less self-regulation is needed for doing relatively less difficult tasks.

MATERIALS AND METHODS

Participants

Participants were 60 German[1] (41 females, age: M = 21.15 years, SD = 1.15) and 62 Iranian (28 females, age: M = 20.53 years, SD = 1.18) undergraduate students. The German participants were recruited from the University of Tübingen in south Germany and Iranian participants were from the University of Tehran, Iran. All participants were native speakers with no immigration backgrounds. The entire data of the participants were analyzed anonymized (i.e., using personal codes instead of names). Detailed characteristics of both German and Iranian students are depicted in **Table 1**.

Measures

Background Characteristics

Background characteristics, consisting of field of study, math score in the University entrance exam, math self-concept, expectancy of success, and demographics of the participants (gender, age, nationality, citizenship, mother tongue, language spoken at home) were collected with a background questionnaire. The questions of the background questionnaire, except the questions of math self-concept, were developed by the authors. Math self-concept was assessed by four questions (e.g., "I am good at mathematics.") based on the SDQ (Self Description Questionnaire) III (Marsh, 1992; German translation: Schwanzer et al., 2005).

Self-Regulation

Participants' self-regulation was assessed by using self-reports. Participants were asked to fill out the Brief Self-Control Scale (BSCS; Tangney et al., 2004; German translation: Bertrams and Dickhäuser, 2009). The German translation of the BSCS (Bertrams and Dickhäuser, 2009) was used in Germany. The original English version of the BSCS was translated into Farsi by two bilingual Ph.D. students from the Psychology field and one bilingual Ph.D. student from outside the field using a well-established method of forward- and backward-translations, following the guidelines from the World Health Organization, 2015).

The BSCS consists of 13 items targeting thought control, impulsive response control, action persistence, and action monitoring (e.g., "I wish I had more self-discipline."). The response format was a 5-point Likert-type scale ranging from 1 (*completely true)* to 5 (*completely untrue*). Nine items were reverse-coded and the total score was the sum of the responses of all items, with higher sum scores representing more self-regulation. In the present study, the questionnaire showed sufficient internal consistency (in German students: Cronbach's $\alpha = 0.84$; in Iranian students: Cronbach's $\alpha = 0.70$).

Mathematics Performance

Mathematics performance was assessed by using the complex multiplication test, consisting of 48 complex multiplication problems. The complex multiplication problems entailed one-digit times two-digit problems with two-digit solutions (e.g., $4 \times 19 = 76$; for further details, see Nemati et al., 2017). The complex multiplication problems and their solutions were presented in a computerized verification task, programmed with the PsychoPy software (Peirce, 2009). Half of the presented solutions were correct, and the other half were incorrect. The task started with eight practice trials. All trials were presented in the center of the screen in a fixed order. The problems and their solutions were presented at the same time after the 500 ms fixation point and remained on the screen until a response was given by the participant, or 6000 ms had passed. Participants responded by pressing the green or red keys (*L* and *A* on a German keyboard) for correct and incorrect solutions, respectively. The response keys were counterbalanced across participants. Except for practice trials, all trials were presented without feedback.

Procedure

All German participants were recruited through e-mail to students and staff of the University of Tübingen and in-person

[1]This study used part of the data of the German participants that was published by Nemati et al. (2017).

TABLE 1 | Descriptive and test statistics of background characteristics and study measurements.

Variable	German			Iranian			Diff
	n	*M* (*SD*)	*K-S*[a]	*n*	*M* (*SD*)	*K-S*[a]	*P*
Age (years)	60	21.15 (1.15)	**<0.001**	62	20.53 (1.18)	**<0.001**	**0.005**[b]
Gender, female	41			28			**0.011**[c]
Field of study							
Human Sciences	40			32			
Engineering/Informatics	20			30			
Math self-concept	60	2.72 (0.80)	**<0.001**	62	2.58 (1.25)	**<0.001**	0.738[b]
Expectancy of success	60	2.88 (0.64)	**<0.001**	62	3.26 (0.92)	**<0.001**	**0.001**[b]
Self-regulation	60	40.92 (8.53)	200	62	42.56 (6.40)	200	0.229[d]
Multiplication performance							
ER	60	0.18 (0.10)	0.077	62	0.19 (0.10)	**0.001**	**<0.001**[b]
RT(s)	60	3.05 (0.54)	200	62	2.59 (0.59)	0.073	**<0.001**[d]

[a]Kolmogorov–Smirnov p-values, [b]Mann–Whitney U Test,[c]Fisher's Exact Test, [d]t-test. Bold p-values depict $p < 0.05$.

contact. All Iranian participants were recruited through flyers and in-person contact in the University of Tehran. The study on German students of Human Sciences was part of a larger project consisting of two testing sessions, each lasting about 2 h, aimed at examining the effects of self-regulatory training on the academic performance of young adults. For their participation, German students of Human Sciences received either course credits or 8 Euro per hour. German students of Engineering/Informatics as well as all the Iranian participants were offered chocolates for their approximately 10 min participation in the study consisting of filling out the background and BSCS questionnaires plus answering the complex multiplication test. First, all participants received detailed information about the study and later gave their written informed consent to participate in the study. The testing session took place in a laboratory in Germany or in an empty classroom of the University of Tehran in Iran. For the variables reported here, each participant was tested individually in a single session. First, all participants were asked to fill out the computerized version of the questionnaires consisting of background questionnaire and BSCS items, which lasted roughly 5 min. Subsequently, they were asked to perform the computerized complex multiplication task, which lasted about 5 min. Participants received a detailed written instruction emphasizing the importance of both speed and accuracy of the responses in the complex multiplication task.

Analysis

Data Preparation

In the present study, better performance in the complex multiplication test was indicated by shorter response times (RTs) and lower error rates (ERs). Multiplication RTs of the participants were defined by the time intervals between the presentation of the multiplication problems on the screen and the responses of the participants, measured by pressing the keys of the computer keyboard. Only RTs of correct responses were considered in the analyses. Moreover, RTs shorter than 200 ms were excluded, and subsequently RTs which were more or less than ± 3 SD around the individual mean were excluded continually until no more outliers remained (see: Nuerk et al., 2001, and follow-up papers for the same method). Accordingly, about 0.1% of the RTs of the German students and 0.2% of the RTs of the Iranian students were excluded. Furthermore, in Germany, two multiplication trials, which were planned to presented with presented with correct solutions, were mistakenly presented with incorrect solutions. Therefore, to keep the match of trials with correct and incorrect solutions, those two trials plus their two equivalent ones with incorrect solutions were excluded from the data of the German students.

Multiplication ERs of the participants were defined as the proportion of incorrect responses. ERs are briefly reported in the descriptive statistics (**Table 1**) but not considered for the further statistical analyses because the performance of German and Iranian students indicated a ceiling effect, as they made few errors in the complex multiplication task (see **Table 1** and **Appendix A**). Finding a ceiling effect in multiplication performance is not surprising as highly educated adults often perform at above-average levels in mathematics tasks (e.g., Siegler and Opfer, 2003; Karolis et al., 2011). Moreover, there were five missing answers in BSCS of two German participants that were replaced by the mean of BSCS answers of the same participants.

Relation of Self-Regulation and Multiplication Performance in German and Iranian Students

The first hypothesis of the present research was that the relation of self-regulation and mathematics performance differs in Germany and Iran. First, to test the effect of self-regulation on multiplication performance in German and Iranian students, a separate linear regression analysis was conducted for each subsample (i.e., German students, Iranian students) with self-regulation as predictor and mean multiplication RTs as outcome variable. In the second step, to compare the relation of self-regulation and mathematics performance between German and Iranian students, the linear regression analysis was calculated with self-regulation, country (dummy coded), and the interaction between self-regulation and country as predictors and mean multiplication RTs as the outcome variable.

Effect of Field of Study on the Relation of Self-Regulation and Multiplication Performance

The second hypothesis of the present research was that the relation of self-regulation and mathematics performance is weaker in students of Engineering/Informatics. In the first step, four separate linear regression analyses were conducted for each subsample field of study (i.e., German and Iranian students of Human Sciences and Engineering/Informatics) with self-regulation as predictor and mean multiplication RTs as outcome variable. In the second step, to compare the relation of self-regulation and mathematics performance in students of Human Sciences and Engineering/Informatics, the interaction between self-regulation and field of study was tested in a multiple linear regression analysis with self-regulation, field of study (dummy coded), and the interaction between self-regulation and field of study as predictors and mean multiplication RTs as the outcome variable. All continuous variables were standardized and the level of significance was set to $\alpha < 0.05$ for all analyses.

RESULTS

Descriptive Statistics

Descriptive and test statistics for the background characteristics and the study measurements of German and Iranian students are presented in **Table 1**[2]. In case of non-normally distributed variables (Kolmogorov–Smirnov test *p*-values < 0.05), Mann–Whitney *U* test, and for normally distributed variables *t*-test and Fisher's Exact Test were used.

German and Iranian students did differ in most of the background characteristics, such as age, $U = 2.40$, $p = 0.005$, gender, $p = 0.011$, Fisher's Exact Test, and expectancy of success, $U = 1.25$, $p = 0.001$. Although German and Iranian students significantly differed in age and gender (see **Table 1**), our result was not explained neither by age nor by gender differences between the two countries (see **Appendix D**).

Additionally, German and Iranian students did differ in their multiplication performance: German students were slower, $t(120) = -4.46$, $p < 0.001$, $d = 0.81$, and Iranian students made more errors, $U = 2.62$, $p < 0.001$. However, German and Iranian students did not differ in math self-concept, $U = 1.92$, $p = 0.738$, and self-regulation, $t(120) = 1.21$, $p = 0.229$, $d = 0.22$.

Relation of Self-Regulation and Multiplication RT in German and Iranian Students

Regression analysis revealed that self-regulation did not predict multiplication RT neither in German ($b = -0.25$, $t = -1.10$, $p = 0.051$; see **Table 2**) nor in Iranian students ($b = -0.09$, $t = -0.72$, $p = 0.473$; see **Table 2**). Moreover, the non-significant interaction indicates that the relation of self-regulation and mathematics performance did not significantly differ between German and Iranian students ($b = -0.09$, $t = -0.53$, $p = 0.599$; see **Table 2**). The data met the assumptions of collinearity (self-regulation, tolerance = 0.36, *VIF* = 2.75*;* country, tolerance = 0.99, *VIF* = 1.01; self-regulation × country, tolerance = 0.37, *VIF* = 2.73*)*, independent errors (Durbin–Watson value = 1.70), and non-zero variances (self-regulation, variance = 56.80; country, variance = 0.25; self-regulation × country, variance = 0.63) and contained no outliers (*Std. Residual Min* = −2.57, *Std. Residual Max* = 2.71).

The Effect of Field of Study on the Relation of Self-Regulation and Multiplication RT

As shown in **Table 3** and **Figure 1**, there is a significant negative relationship between self-regulation and multiplication RT in German [$b = -0.35$, $t = -2.26$, $p = 0.029$; Model 1: $R^2 = 0.12$, $F(1,38) = 5.12$, $p = 0.029$] and Iranian [$b = -0.29$, $t = -2.23$, $p = 0.034$; Model 3: $R^2 = 0.14$, $F(1,30) = 4.96$, $p = 0.034$] students of Human Sciences, but not in German ($b = -0.07$, $t = -0.34$, $p = 0.736$) and Iranian ($b = 0.12$, $t = 0.80$, $p = 0.428$) students of Engineering/Informatics (see **Table 3**). Similar decreasing trends in Human Sciences showed in **Figure 1**, indicating the higher the self-regulation the better the students of Human Sciences within each countries performed in the complex multiplication task.

Moreover, the non-significant interaction indicates that the relation of self-regulation and mathematics performance did not significantly differ between students of Human Sciences and Engineering/Informatics ($b = 021$, $t = 131$, $p = 0.194$; **Table 3**). The data met the assumptions of collinearity *(*self-regulation, tolerance = 0.63*, VIF* = 1.60; field of study, tolerance = 0.99, *VIF* = 1.00; self-regulation × field of study, tolerance = 0.63, *VIF* = 1.60), independent errors (Durbin–Watson value = 1.83), and non-zero variances (self-regulation, variance = 56.80; field of study, variance = 0.24; self-regulation × field of study, variance = 0.37) and contained no outliers (*Std. Residual Min* = −2.49, *Std. Residual Max* = 2.35).

TABLE 2 | Regression analysis predicting multiplication RT from self-regulation in German and Iranian students.

Predictor	b	SE(B)	t	p
Model 1				
Self-regulation in Germans[a]	−0.25	0.13	−1.10	0.051
Model 2				
Self-regulation in Iranians[b]	−0.09	0.13	−0.72	0.473
Model 3				
Constant	−0.36	0.12	−3.03	**0.003**
Self-regulation	−0.11	0.14	−0.76	0.446
Country	0.72	0.17	4.27	**<0.001**
Self-regulation × country	−0.09	0.17	−0.53	0.599

[a]n = 60, [b]n = 62. All variables are standardized and country was dummy coded. Bold p-values depict p < 0.05.

[2]Descriptive statistics for each subsample field of study (i.e., German and Iranian students of Human Sciences and Engineering\Informatics) and the correlation matrix in German and Iranian students are presented in **Appendices B**, **C**, respectively for the interested reader.

TABLE 3 | Regression analysis predicting multiplication RT from self-regulation in German and Iranian students of Human Sciences and Engineering/Informatics.

Predictor	*b*	*SE(B)*	*t*	*p*
Model 1				
Self-regulation of German students of Human Sciences[a]	−0.35	0.15	−2.26	**0.029**
Model 2				
Self-regulation of German students of Engineering/Informatics[b]	−0.07	0.22	−0.34	0.736
Model 3				
Self-regulation of Iranian students of Human Sciences[c]	−0.29	0.13	−2.23	**0.034**
Model 4				
Self-regulation of Iranian students of Engineering/Informatics[d]	0.12	0.13	0.80	0.428
Model 5				
Constant	0.38	0.10	3.75	**<0.001**
Self-regulation	−0.30	0.10	−3.01	**0.003**
Field of study	−0.93	0.16	−5.79	**<0.001**
Self-regulation × field of study	0.21	0.16	1.31	0.194

[a]n = 40, [b]n = 20, [c]n = 34, [d]n = 30. All variables are standardized and country was dummy coded. Bold p-values depict $p < 0.05$.

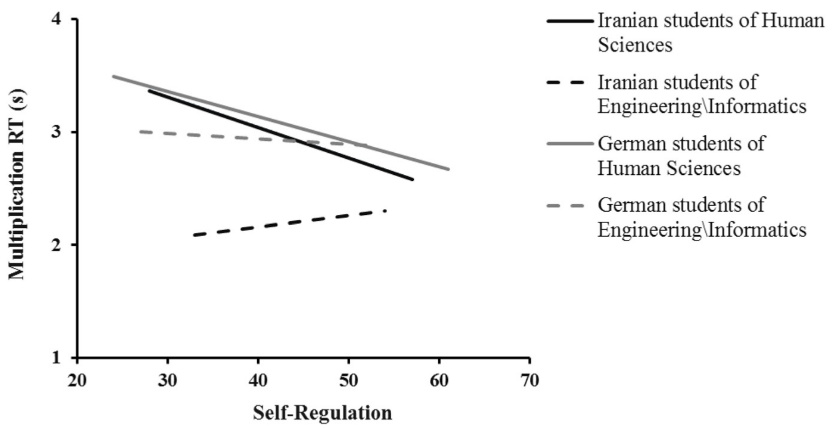


FIGURE 1 | Linear regression trend lines testing the relation of self-regulation as predictor and multiplication RT as the dependent variable in four different subsamples: German students of Human Sciences, German students of Engineering/Informatics, Iranian students of Human Sciences, and Iranian students of Engineering/Informatics.

DISCUSSION

The present study investigated whether the relation of self-regulation and mathematics performance differs between students in two different contexts, namely independent and interdependent cultures (i.e., Germany vs. Iran). As the second hypothesis, we expected that the relation of self-regulation and mathematics performance was weaker in students of Engineering/Informatics as compared to students of Human Sciences. Contradictory to our first hypothesis, the relation of self-regulation and mathematics performance did not differ between German and Iranian college students: self-regulation did not predict multiplication RT neither in German nor Iranian students. Moreover, inconsistent with our second hypothesis, the results showed that the relation of self-regulation and mathematics performance did not differ significantly between students studying less math-related fields (i.e., Human Sciences) and students of math-related fields (i.e., Engineering/Informatics) in the whole sample. However, partially in line with our second hypothesis, when the field of study was considered within the countries, self-regulation predicted multiplication RT in those students studying Human Sciences but not in students of Engineering/Informatics within each country. Thus, although the main effect of field of study was not observed regardless of country, the relation of self-regulation and mathematics performance seemed to be descriptively weaker in students of Engineering/Informatics than Human Sciences within each country. This might be because the complex multiplication test within each country seemed to be less difficult for the students of Engineering/Informatics compared to the students of Human Sciences, therefore, these students might need less self-regulation to solve the problems. The complex multiplication test seemed to be less difficult for the students of Engineering/Informatics as they performed better (i.e., they had shorter RT and less ER) than students of Human Sciences in general (see **Appendix B**). However, this effect was significantly different in Iranians but only descriptively different in Germans (see **Appendix B**). Moreover, expectancy of success in math was higher in students of Engineering/Informatics than students of Human Sciences (see **Appendix B**) within each country and significantly correlated with shorter RTs (see **Appendix C**), suggesting that students of Engineering/Informatics believed in their self-ability to do well in mathematics. Thus, students of Engineering/Informatics within each country might have used less self-regulation while doing complex multiplication test as the test was less difficult for them. This is consistent with previous studies revealing that individuals need more self-regulation while solving challenging tasks (e.g., Ackerman, 1989; Kanfer and Ackerman, 1989; Steele-Johnson et al., 2000). For instance, it has been shown that task difficulty can moderate the effect of self-regulation on performance (Steele-Johnson et al., 2000). The authors found that when the cognitive load of the task is high, individuals have to decide how to allocate their limited attentional resources to the task, therefore, they are in need of more self-regulation.

Taken together, the results showed that the relation of self-regulation and mathematics performance did not differ between German and Iranian college students. Furthermore, we observed this similarity not only in the context of country but also in the context of field of study, which is further supported by the fact that when only the students of Human Sciences are compared, the association between self-regulation and mathematics is similar in both countries (**Appendix E**). This finding is in line with a cross-cultural study by Lan et al. (2011), described earlier, that assessed the cognitive components of self-regulation, such as inhibition and attentional control, and examined their associations with simple and complex mathematics performances in Chinese and North American children. Their results demonstrated that the relation of different cognitive aspects of self-regulation and both simple and complex mathematics performance are similar in Chinese and North American children. The authors argued that the neurobiological and genetic factors which determine the strength of associations between various components of self-regulation may be similar in distinct contexts, therefore, their subsequent contribution to academic performance is also more likely to be consistent across countries. However, Chinese children outperformed North American children in some of the self-regulation tasks such as inhibition and attentional control. The authors ascribed these performance differences in self-regulation tasks to variances in specific cultural practices in educational settings during kindergarten and primary school. For instance, it has been shown that Asian children receive more intensive practice in controlling their attention and behavior in kindergarten or the classroom than North American children (e.g., Chen et al., 1998; Kwon, 2004; Lan et al., 2009). Therefore, it seems that although different aspects of self-regulation may be learned and used differently in interdependent and independent countries, their interrelations with each other and their association with mathematics performance remains similar. This interpretation is also in line with the idea that both independent and interdependent systems exist and are essential in each country, but there might be differences among the countries in the strength of their application (e.g., Harwood et al., 2001; Leyendecker et al., 2002; Jing-Schmidt, 2014). In the same vein, both independent and interdependent self-regulation processes may exist in Germany and Iran to different degrees, but this may not significantly influence their level of contribution to the mathematics performance.

However, our finding is in contrast with previous studies, connecting the academic achievement gap between students from different countries to the effect of cultural context on self-regulation. For instance, in a longitudinal study by Nota et al. (2004) which was explained earlier, many of the self-regulatory strategies that predicted academic achievement in American students did not directly predict academic achievement in Italian students. However, compared to the study by Nota et al. (2004), the effect of various self-regulatory strategies was not investigated in the present study and contexts as well as measures of self-regulation and academic achievement differ from their study. Another important reason why, unlike our study, they found differences in the relation of self-regulation and academic achievement across two countries, might be the effect of samples: Italian students were high achievers who are more likely to self-regulate than typical populations of students and in this sense differed from the American students or from German and Iranian students in our research.

Altogether, cultural context did not seem to play a dominant role in moderating the relation between self-regulation and math performance in the present study. However, with regard to the confounding effect of field of study within each country on the predictive validity of self-regulation, careful sample selection considering field of study of students is recommended for future research examining the relation of self-regulation and mathematics performance.

Limitations

The current research has some limitations worth noting. First, there might be structural and cultural variations in educational systems such as different grading systems or teachers' expectations, as well as academic motivation of students within and between nations that may differentially influence self-regulation and its relation with academic performance. Therefore, we view this study only as a starting point for investigating the impact of independent and interdependent cultures on the relation of self-regulation and math performance. Future studies conducted in other independent or interdependent cultures should clarify whether the observed results are really due to this cultural difference or to other educational or cultural differences, which are particular to the specific countries studied here. Second, German students of Human Sciences were offered different reimbursement than other participants since the study in which they participated, was part of a larger project consisting of 4-h experiment. Hence, we acknowledge that different incentives in German students of Human Sciences in comparison to other participants might generate participation bias and account partially for the findings of the current study. Third, self-regulation consists of several components such as cognitive, behavioral, and emotional aspects that are differentially related to mathematics performance and their effects should be investigated individually in the future research. Forth limitation is the small sample size of the present study that may preclude a definitive statement for the present study. The last, but not least, important limitation is construct validity in the present study, as our research measurement for assessing self-regulation was designed and validated for Western countries. The problem is that in self-reports, participants of one cultural context may interpret the words differently and compare themselves with different standards than those in another cultural context (e.g., Heine et al., 2002). In our study, the internal consistency of the self-regulation self-report in Iranian students is sufficient for the present study and in line with previously reported findings in Eastern countries such as China (Cronbach's $\alpha = 0.75$; Unger et al., 2016), however, it should be also noted that it is relatively low, which can be due to either a reliability or homogeneity problem. In the future, international researchers should strive for a transcultural self-regulation scale, which can be used in Western and non-Western cultures with high reliability and validity.

CONCLUSION

In conclusion, our findings show that the relation of self-regulation and mathematics performance is similar in German and Iranian college students. In addition, the effect of field of study on the relation of self-regulation and mathematics performance was highlighted in the present study. Self-regulation did not predict mathematics performance in German and Iranian students, however, when the effect of field of study was taken into account, self-regulation predicted mathematics performance in students of less math-related fields of study within each country. It is important to note that while the single analysis produced differential results, a direct comparison of the different fields of studies was non-significant – therefore, we have interpreted these results with great care. Nevertheless, since the relation between self-regulation and mathematics performance within each country, was significant only for less math-related fields of study, we suggest that the possible confounding effect of field of study should be considered in studies when the relation of self-regulation and mathematics performance is examined.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PN and CG designed and performed the research. PN and JK analyzed the data. PN, JK, CG, and H-CN wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was financed by money from the chair school psychology of the University of Tübingen.

ACKNOWLEDGMENTS

We would like to thank all participating students. We would also like to thank our research assistants: Merle Bode for her help in recruiting participants in Germany and Zoe Lauren Kirste for supporting language proofreading of the manuscript.

REFERENCES

Ackerman, P. L. (1989). "Individual differences in skill acquisition," in *Learning and Individual Differences: Advances in Theory and Research*, eds P. L. Ackerman, R. J. Sternberg, and R. Glaser (New York, NY: Freeman), 165–217.

Baumeister, R. F., and Vohs, K. D. (2007). Self-regulation, ego depletion, and motivation. *Soc. Pers. Psychol. Comp.* 1, 115–128. doi: 10.1111/j.1751-9004.2007.00001.x

Bertrams, A., and Dickhäuser, O. (2009). Messung dispositioneller Selbstkontroll-Kapazität: Eine deutsche Adaptation der Kurzform der Self-Control Scale (SCS-KD). *Diagnostica* 55, 2–10. doi: 10.1026/0012-1924.55.1.2

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

Blair, C., and Ursache, A. (2011). "A bidirectional model of executive functions and selfregulation," in *Handbook of Selfregulation: Research, Theory and Applications*, 2nd Edn, eds R. F. Baumeister and K. D. Vohs (New York, NY: Guilford Press), 300–320.

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

Camahalan, F. M. G. (2006). Effects of self-regulated learning on mathematics achievement of selected Southeast Asian children. *J. Instructional Psychol.* 33, 194–206.

Carlson, V. J., and Harwood, R. L. (2003). Attachment, culture and the caregiving system: the cultural patterning of everyday experiences among Anglo and Puerto Rican mother-infant pairs. *Infant Mental Health J.* 24, 53–73. doi: 10.1002/imhj.10043

Chen, X., Hastings, P. D., Rubin, K. H., Chen, H., Cen, G., and Stewart, S. L. (1998). Child-rearing attitudes and behavioral inhibition in Chinese and Canadian toddlers: a cross-cultural study. *Dev. Psychol.* 34, 677–686. doi: 10.1037/0012-1649.34.4.677

Domahs, F., Delazer, M., and Nuerk, H.-C. (2006). What makes multiplication facts difficult: Problem size or neighbourhood consistency? *Exp. Psychol.* 53, 275–282. doi: 10.1027/1618-3169.53.4.275

Domahs, F., Janssen, U., Schlesewsky, M., Ratinckx, E., Verguts, T., Willmes, K., et al. (2007). Neighborhood consistency in mental arithmetic: Behavioral and ERP evidence. *Behav. Brain Funct.* 3:66. doi: 10.1186/1744-9081-3-66

Ellefson, M., Ng, F. F., Wang, Q., and Hughes, C. (2017). Efficiency of executive function: a two-generation cross-cultural comparison of samples from Hong Kong and the United Kingdom. *Psychol. Sci.* 28, 555–566. doi: 10.1177/0956797616687812

Frazier, T. W., Youngstrom, E. A., Glutting, J. J., and Watkins, M. W. (2007). ADHD and achievement: meta-analysis of the child, adolescent, and adult literatures and a concomitant study with college students. *J. Learn. Disabil.* 40, 49–65. doi: 10.1177/00222194070400010401

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *J. Educ. Psychol.* 98, 29–43. doi: 10.1037/0022-0663.98.1.29

Gawrilow, C., Fäsche, A., Guderjahn, L., Gunzenhauser, C., Merkt, J., and von Suchodoletz, A. (2014). The impact of self-regulation on preschool mathematical achievement. *Child Indicators Res.* 7, 805–820. doi: 10.1007/s12187-013-9201-y

Gawrilow, C., Gollwitzer, P. M., and Oettingen, G. (2011). If-then plans benefit executive functions in children with ADHD. *J. Soc. Clin. Psychol.* 30, 616–646. doi: 10.1521/jscp.2011.30.6.616

Harwood, R., Penn-Handwerker, W., Schoelmerich, A., and Leyendecker, B. (2001). Ethnic category labels, parental beliefs and the contextualized individual: an exploration of the individualism-sociocentrism debate. *Parent. Sci. Pract.* 1, 217–236. doi: 10.1207/S15327922PAR0103_03

Heine, S. J., Lehman, D. R., Peng, K., and Greenholtz, J. (2002). What's wrong with cross-cultural comparisons of subjective Likert scales: The reference-group effect. *J. Pers. Soc. Psychol.* 82, 903–918. doi: 10.1037/0022-3514.82.6.903

Higgins, E. T., Pierro, A., and Kruglanski, A. W. (2008). "Re-thinking culture and personality: how self-regulatory universals create cross-cultural differences," in *Motivation and Cognition Across Cultures*, eds R. M. Sorrentino and S. Yamaguchi (Amsterdam: Academic), 161–190. doi: 10.1016/B978-0-12-373694-9.00008-8

Hofmann, W., Friese, M., Schmeichel, B. J., and Baddeley, A. D. (2011). "Working memory and self-regulation," in *Handbook of Self-regulation: Research, Theory, and Applications*, 2nd Edn, eds K. D. Vohs and R. F. Baumeister (New York, NY: The Guilford Press), 204–225.

Hofstede, G. (1980). *Culture's Consequences*. Beverly Hills, CA: Sage Publications.

Jaramillo, J. M., Rendón, M. I, Muñoz, L., Weis, M., and Trommsdorff, G. (2017). Children's self-regulation in cultural contexts: the role of parental socialization theories, goals, and practices. *Front. Psychol.* 8:923. doi: 10.3389/fpsyg.2017.00923

Jing-Schmidt, Z. (2014). "Maternal affective input in mother-child interaction. A cross-cultural perspective," in *Development of Pragmatic and Discursive Skills in Chinese-Speaking Children*, eds Z. Hua and L. Jin (Amsterdam: John Benjamins Publishing), 57–90. doi: 10.1075/bct.60.04jin

Kanfer, R., and Ackerman, P. L. (1989). Motivation and cognitive abilities: an integrative/aptitude-treatment interaction approach to skill acquisition. *J. Appl. Psychol.* 74, 657–690. doi: 10.1037/0021-9010.74.4.657

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

Keller, H., Yovsi, R., Borke, J., Kärtner, J., Jensen, H., and Papaligoura, Z. (2004). Developmental consequences of early parenting experiences: self-recognition and self-regulation in three cultural communities. *Child Dev.* 75, 1745–1760. doi: 10.1111/j.1467-8624.2004.00814.x

Kwon, Y.-I. (2004). Early childhood education in Korea: discrepancy between national kindergarten curriculum and practices. *Educ. Rev.* 56, 297–312. doi: 10.1080/0013191042000201208

Labuhn, A. S., Zimmerman, B. J., and Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: the influence of feedback and self-evaluative standards. *Metacogn. Learn.* 5, 173–194. doi: 10.1007/s11409-010-9056-2

Lamm, B., Keller, H., Teiser, J., Gudi, H., Yovsi, R. D., Freitag, C., et al. (2018). Waiting for the second treat: developing culture−specific modes of self−regulation. *Child Dev.* 89, 261–277. doi: 10.1111/cdev.12847

Lan, X., Legare, C. H., Ponitz, C. C., Li, S., and Morrison, F. J. (2011). Investigating the links between the subcomponents of executive function and academic achievement: a cross-cultural analysis of Chinese and American preschoolers. *J. Exp. Child Psychol.* 108, 677–692. doi: /10.1016/j.jecp.2010.11.001

Lan, X., Ponitz, C. C., Miller, K. F., Li, S., Cortina, K., Perry, M., et al. (2009). Keeping their attention: classroom practices associated with behavioral engagement in first grade mathematics classes in China and the United States. *Early Childhood Res. Q.* 24, 198–211. doi: 10.1016/j.ecresq.2009.03.002

Leyendecker, B., Lamb, M. E., Harwood, R., and Schölmerich, A. (2002). Mothers' socialization goals and evaluations of desirable and undesirable everyday situations in two diverse cultural groups. *Int. J. Behav. Dev.* 26, 248–258. doi: 10.1080/01650250143000030

Markus, H., and Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion and motivation. *Psychol. Rev.* 98, 224–253. doi: 10.1037/0033-295X.98.2.224

Marsh, H. W. (1992). *Self Description Questionnaire (SDQ) III: A Theoretical and Empirical basis for the Measurement of Multiple Dimensions of late Adolescent Self-concept: an Interim Test Manual and a Research Monograph*. Macarthur, NSW: University of Western Sydney, Faculty of Education.

McClelland, M. M., and Cameron, C. E. (2011). "Self-regulation and academic achievement in elementary school children," in *Thriving in Childhood and Adolescence: The Role of Self-regulation Processes. New Directions for Child and Adolescent Development*, Vol. 3, eds R. M. Lerner, J. V. Lerner, E. P. Bowers, S. Lewin-Bizan, S. Gestsdottir, and J. B. Urban (Hoboken, NJ: Wiley), 29–44.

McClelland, M. M., Ponitz, C. C., Messersmith, E., and Tominey, S. (2010). "Self-regulation: integration of cognition and emotion," in *The Handbook of Life-Span Development: Cognition, Biology and Methods*, Vol. 1, eds R. Lerner and W. Overton (Hoboken, NJ: Wiley & Sons), 509–553. doi: 10.1002/9780470880166.hlsd001015

Nemati, P., Schmid, J., Soltanlou, M., Krimly, J. T., Nuerk, H. C., and Gawrilow, C. (2017). Planning and self-control, but not working memory, directly predict multiplication performance in adults. *J. Num. Cogn.* 3, 441–467. doi: 10.5964/jnc.v3i2.61

Nota, L., Soresi, S., and Zimmerman, B. J. (2004). Self-regulation and academic achievement and resilience: a longitudinal study. *Int. J. Educ. Res.* 41, 198–215. doi: 10.1016/j.ijer.2005.07.001

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

Otts, D. C. (2010). Self-Regulation and Math Attitudes: Effects on Academic Performance In Developmental Math Courses at a Community College. Doctor Dissertation, University of Kansas, Kansas.

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform.* 2:10. doi: 10.3389/neuro.11.010.2008

Rao, N., Moely, B. E., and Sachs, J. (2000). Motivational beliefs, study strategies, and mathematics attainment in high-and low-achieving Chinese secondary school students. *Contemp. Educ. Psychol.* 25, 287–316. doi: 10.1006/ceps.1999.1003

Schwanzer, A. D., Trautwein, U., Lüdtke, O., and Sydow, H. (2005). Entwicklung eines Instruments zur Erfassung des Selbstkonzepts junger Erwachsener. *Diagnostica* 51, 183–194. doi: 10.1026/0012-1924.51.4.183

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

Steele-Johnson, D., Beauregard, R. S., Hoover, P. B., and Schmidt, A. M. (2000). Goal orientation and task demand effects on motivation, affect, and performance. *J. Appl. Psychol.* 85:724. doi: 10.1037/0021-9010.85.5.724

Tangney, J. P., Baumeister, R. F., and Boone, A. L. (2004). High self−control predicts good adjustment, less pathology, better grades, and interpersonal success. *J. Pers.* 72, 271–324. doi: 10.1111/j.0022-3506.2004.00263.x

Trommsdorff, G. (2009). Culture and development of self-regulation. *Soc. Pers. Psychol. Compass* 3, 687–701. doi: 10.1111/j.1751-9004.2009.00209.x

Unger, A., Bi, C., Xiao, Y. Y., and Ybarra, O. (2016). The revising of the tangney self−control Scale for Chinese students. *PsyCh J.* 5, 101–116. doi: 10.1002/pchj.128

von Suchodoletz, A., and Gunzenhauser, C. (2013). Behavior regulation and early math and vocabulary knowledge in German preschool children. *Early Educ. Dev.* 24, 310–331. doi: 10.1080/10409289.2012.693428

von Suchodoletz, A., Uka, F., and Larsen, R. A. (2015). Self-regulation across different contexts: findings in young Albanian children. *Early Educ. Dev.* 26, 829–846. doi: 10.1080/10409289.2015.1012189

World Health Organization (2015). *Process of Translation and Adaptation of Instruments*. Available online at: http://www.who.int/substance_abuse/research_tools/translation/en/index.html (accessed April 15, 2015).

Zentall, S. S. (2007). "Math performance of students with ADHD: Cognitive and behavioral contributors and interventions," 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: Paul H Brookes Publishing), 219–243.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: an overview. *Educ. Psychol.* 25, 3–17. doi: 10.1207/s15326985ep2501_2

Conflict of Interest: 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 A

Ceiling/Floor Effect for Multiplication ER

Sample	*n*	*M (SD)*	*K-S*[a]	*Skewness*	*Kurtosis*
German students	60	0.18 (0.10)	0.077	0.94	1.53
Iranian students	62	0.19 (0.10)	0.001	1.16	1.36

[a]*Kolmogorov–Smirnov p-values.*

APPENDIX B

Descriptive and Test Statistics of Background Characteristics and Study Measurements in German and Iranian Students of Human Sciences and Engineering/Informatics

TABLE B1 | Analysis of variance of country and field of study on ER.

Source	*df*	*MS*	*F*	*p*	Partial η^2
Country	1	0.14	18.64	**<0.001**	0.14
Field of study	1	0.13	18.52	**<0.001**	0.14
Country × field of study	1	0.13	17.76	**<0.001**	0.13
Error	118	0.01			

MS, mean squares. Bold p-values depict $p < 0.05$.

TABLE B2 | Analysis of variance of country and field of study on RT.

Source	*df*	*MS*	*F*	*p*	Partial η^2
Country	1	5.61	23.01	**<0.001**	0.16
Field of study	1	6.47	26.60	**<0.001**	0.18
Country × field of study	1	2.71	11.12	**0.001**	0.09
Error	118	0.24			

MS, mean squares. Bold p-values depict $p < 0.05$.

TABLE B3 | Descriptive of German students of Human Sciences and Engineering/Informatics.

Variable	Human Sciences		Engineering/Informatics		Diff
	M (SD)	*K-S*[a]	*M (SD)*	*K-S*[a]	*p*
Age (years)	20.95 (1.08)	0.001	21.55 (1.19)	<0.001	**0.039**[b]
Gender, female (%)	82		40		**0.002**[c]
Math self-concept	2.75 (0.84)	<0.001	2.65 (0.74)	<0.001	0.568[b]
Expectancy of success	2.78 (0.73)	<0.001	3.10 (0.31)	<0.001	0.052[b]
Self-regulation	41.43 (8.74)	0.200	39.90 (8.20)	0.200	0.518[d]
Multiplication performance					
ER	0.19 (0.11)	0.028	0.18 (0.08)	0.200	0.742[b]
RT(s)	3.10 (0.56)	0.200	2.93 (0.48)	0.200	0.259[d]

[a]*Kolmogorov–Smirnov p-values,* [b]*Mann–Whitney U test,*[c]*Fisher's Exact Test,* [d]*t-test. Bold p-values depict $p < 0.05$.*

TABLE B4 | Descriptive of Iranian students of Human Sciences and Engineering/Informatics.

Variable	Human Sciences		Engineering/Informatics		Diff
	M (*SD*)	*K-S*[a]	*M* (*SD*)	*K-S*[a]	*p*
Age (years)	20.09 (1.20)	<0.001	21.00 (0.98)	0.004	**0.001**[b]
Gender, female (%)	50		40		0.456[c]
Math self-concept	1.91 (1.28)	<0.001	3.03 (0.70)	<0.001	**<0.001**[b]
Expectancy of success	2.88 (1.01)	<0.001	3.67 (0.37)	<0.001	**<0.001**[b]
Self-regulation	42.69 (6.57)	0.200	42.43 (6.33)	0.200	0.877[d]
Multiplication performance					
ER	0.18 (0.09)	0.151	0.05 (0.04)	<0.001	**<0.001**[b]
RT(s)	2.97 (0.47)	0.007	2.18 (0.42)	0.200	**<0.001**[b]

[a]Kolmogorov–Smirnov p-values, [b]Mann–Whitney U test,[C]Fisher's Exact Test, [d]t-test. Bold p-values depict $p < 0.05$.

APPENDIX C

Correlation Matrix in German and Iranian Students

TABLE C1 | Correlations between background variables, self-regulation, and multiplication RT in German students.

Variable	1	2	3	4	5	6
1. Age	–					
2. Gender	0.29*	–				
3. Math self-concept	−0.21	0.20	–			
4. Expectancy of success	−0.21	0.24	0.43*	–		
5. Self-regulation	−0.14	−0.26*	−0.05	0.07	–	
6. Multiplication RT(s)	0.01	−0.18	−0.11	−0.28*	−0.25	–

*n = 60. *Correlation is significant at the 0.05 level (2-tailed).*

TABLE C2 | Correlations between background variables, self-regulation, and multiplication RT in Iranian students.

Variable	1	2	3	4	5	6
1. Age	–					
2. Gender	0.22	–				
3. Math self-concept	0.29*	0.14	–			
4. Expectancy of success	0.13	0.01	0.34*	–		
5. Self-regulation	0.12	0.10	0.13	−0.17	–	
6. Multiplication RT(s)	−0.11	−0.03	−0.61*	−0.58*	−0.09	–

*n = 62. *Correlation is significant at the 0.05 level (2-tailed).*

APPENDIX D

Linear Model of Age and Gender as Predictors of Multiplication RT

TABLE D1 | Linear model of predictors of multiplication RT.

Predictor	*b*	*SE(B)*	*t*	*p*
Constant	−0.00	0.09	−0.04	0.964
Self-regulation	−0.19	0.09	−2.16	**0.032**
Age	0.04	0.09	0.41	0.684
Self-regulation × age	−0.08	0.09	−0.90	0.372

N = 122. All variables are standardized. $R^2 = 0.05$, $F(3,118) = 2.05$, $p = 0.111$. *Bold p-values depict* $p < 0.05$.

TABLE D2 | Linear model of predictors of multiplication RT.

Predictor	*b*	*SE(B)*	*t*	*p*
Constant	−0.40	0.12	−3.23	**0.002**
Age	−0.10	0.12	−0.86	0.389
Country	0.78	0.17	4.43	**0.000**
Age × country	0.11	0.18	0.63	0.527

N = 122. All variables are standardized and country was dummy coded. $R^2 = 0.15$, $F(3,118) = 6.81$, $p < 0.001$. *Bold p-values depict* $p < 0.05$.

TABLE D3 | Linear model of predictors of multiplication RT.

Predictor	*b*	*SE(B)*	*t*	*p*
Constant	0.17	0.12	1.48	0.142
Self-regulation	−0.28	0.12	−2.39	0.019
Gender	−0.39	0.18	−2.18	0.031
Self-regulation × gender	0.14	0.18	0.81	0.419

N = 122. All variables are standardized and Gender was dummy coded. $R^2 = 0.08$, $F(3,118) = 3.61$, $p = 0.015$.

APPENDIX E

Linear Model of Self-Regulation as the Predictors of Multiplication RT in Students of Human Sciences

TABLE E1 | Linear model of predictors of multiplication RT in students of Human Sciences.

Predictor	*b*	*SE(B)*	*t*	*p*
Constant	0.29	0.15	2.01	**0.048**
Self-regulation	−0.33	0.17	−1.98	0.051
Country	0.17	0.19	0.87	0.389
Self-regulation × country	0.06	0.20	0.29	0.773

n = 72. All variables are standardized. $R^2 = 0.14$, $F(3,68) = 3.75$, $p = 0.015$. *Bold p-values depict* $p < 0.05$.

ORIGINAL RESEARCH
published: 04 November 2020
doi: 10.3389/fpsyg.2020.518981

The Multifactorial Nature of Early Numeracy and Its Stability

David Braeuning[1,2,3], Andrew Ribner[4], Korbinian Moeller[1,3,5,6] and Clancy Blair[4] on behalf of Family Life Project Key Investigators*

[1] LEAD Graduate School & Research Network, University of Tübingen, Tübingen, Germany, [2] Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Tübingen, Germany, [3] Leibniz-Institut für Wissensmedien, Tübingen, Germany, [4] Department of Applied Psychology, New York University, New York, NY, United States, [5] Department of Psychology, University of Tübingen, Tübingen, Germany, [6] Centre For Mathematical Cognition, School of Science, Loughborough University, Loughborough, United Kingdom

Early numeracy is a robust predictor of later mathematical abilities. So far, early numeracy has typically been presented as a unitary or two-factorial construct. Nevertheless, there is recent evidence suggesting that it may also be reflected by more basic numerical competences. However, the structure and stability of such a multifactorial model of early numeracy over time has not been investigated yet. In the present study, we used data from a large, longitudinal sample (N = 1292) in the United States with assessments of math ability in prekindergarten and kindergarten to evaluate both the factorial structure of early numeracy and its stability over time. Confirmatory factor analysis identified four distinct basic numerical competences making up early numeracy in prekindergarten: patterning/geometry, number sense, arithmetic, and data analysis/statistics. Stability as tested by means of measurement invariance indicated configural invariance of these four factors from prekindergarten to kindergarten. This reflected that early numeracy in kindergarten was made up by the same four basic numerical competences as in prekindergarten and thus seemed rather stable over the course of preschool. These findings may not only have implications for research on numerical cognition but particularly for diagnostic processes or the development of interventions in educational practice.

Keywords: early numeracy, basic numerical competences, mathematical abilities, structure, stability, predictor

Edited by:
Snehlata Jaswal, Chaudhary Charan Singh University, India

Reviewed by:
Robert Reeve, The University of Melbourne, Australia
Ann Dowker, University of Oxford, United Kingdom

****Correspondence:***
David Braeuning
david.braeuning@uni-tuebingen.de; braeuning.d@web.de

Specialty section:
This article was submitted to Cognitive Science, a section of the journal Frontiers in Psychology

Received: *11 December 2019*
Accepted: *07 October 2020*
Published: *04 November 2020*

Citation:
Braeuning D, Ribner A, Moeller K and Blair C (2020) The Multifactorial Nature of Early Numeracy and Its Stability. Front. Psychol. 11:518981. doi: 10.3389/fpsyg.2020.518981

INTRODUCTION

Basic numerical competences acquired before school-entry are important predictors for later mathematical and educational achievement (e.g., Parsons and Bynner, 2005; Duncan et al., 2007; Jordan et al., 2009, 2010). These competences are often summarized under the broad construct *early numeracy* (e.g., Lembke and Foegen, 2009; Aunio and Niemivirta, 2010). Although it has been suggested that the construct of early numeracy is more accurately represented by multiple distinct basic numerical competences (Dowker, 2008), very few studies have examined the specific structure of basic numerical competences making up early numeracy prior to school entry.

Moreover, previous longitudinal studies have typically investigated whether and—if so—which basic numerical competences predict later mathematical achievement operationalized in terms of scores of (standardized) math tests or sometimes math grades (e.g., Parsons and Bynner, 2005; Jordan et al., 2010). As such, they do not describe the development of basic numerical competences themselves, but how they predict other, usually more complex arithmetical and mathematical abilities. In turn, little is known about the stability of basic numerical competences that make up early numeracy and the ways in which they develop over time. Hence, this study aims to evaluate the specific structure of early numeracy by specifying its underlying basic numerical competences and their stability across the transition from preschool (age 5) to kindergarten (age 6).

In the following, we will first give an overview of uni- and multi-dimensional conceptualizations of early numeracy and its dimensionality. We then review previous findings on the stability of numeracy performance and basic numerical competences.

From a Uni- to a Multi-Dimensional Perspective on Early Numeracy

Previous research on cognitive development has often considered early numeracy as a unidimensional skill. Accordingly, it is subsumed under a single parameter score that reflects performance over a broad range of tasks, covering primarily numerical (e.g., counting, number knowledge, basic calculations; e.g., Jordan et al., 2007, 2009, 2010; Kroesbergen et al., 2009) but also more visual-spatial processes (e.g., recognition of shapes or patterns, geometry; e.g., Jordan et al., 2006; Anders et al., 2012; Polignano and Hojnoski, 2012). However, such a unidimensional conceptualization of early numeracy that averages out contributions of specific basic numerical competences can only provide a rather general measure of early numeracy but not reflect its underlying structure of basic numerical competences adequately. Indeed, Dowker (2008) suggested that children in the preschool years are already capable of performing numerical tasks that require distinct basic competences, suggesting numeracy might be multidimensional even in early childhood prior to formal schooling.

Practically speaking, children develop numerical competences in distinct domains that often correspond to the way that mathematics is taught to them. Content analysis of elementary mathematics textbooks from kindergarten through sixth grade has documented that since the 1960s, mathematics instruction has expanded considerably, in particular in the topics covered (e.g., operations, geometry, patterns, etc.), as well as in the introduction of advanced topics at increasingly earlier grades (Baker et al., 2010). In particular, math education usually differentiates math competences on a conceptual level following content strands (National Council of Teachers of Mathematics, 1989, 2000). These include children's understanding of (i) properties of numbers, as well as arithmetic operations (e.g., addition, multiplication) and their application to real-world situations, (ii) operating with measurement units like money, time, etc., (iii) geometry from shapes to transformations, (iv) data analysis and statistics as reflected in collecting, organizing, reading, and representing data, and (v) recognition of patterns and functions. These groupings of content areas within mathematics education suggest variation in the types of mathematical competences that children acquire and indicate the potential value of examining variation in distinct domains of mathematical competences when drawing conclusions about achievement in elementary grades.

Previous research following a multidimensional perspective has identified different basic numerical competences to make up the construct of early numeracy. Mostly, two-factorial models have been suggested differentiating, for instance, relational abilities and counting (e.g., Aunio et al., 2004), symbolic and non-symbolic numerical abilities (e.g., Kolkman et al., 2013), or procedural and conceptual abilities (e.g., Ribner et al., 2018). However, these models still reflect rather broad descriptions of early numeracy and few studies have examined further and more specific differentiations that might be more aligned with curricular approaches (e.g., Cirino, 2011). For instance, one such study on the structure of early numeracy comes from an analysis of large-scale assessment data from more than 1,700 5- to 6-year-old children in the Netherlands. In this study, Hirsch et al. (2018) categorized items from an early numeracy test according to the distinct basic numerical competences theoretically underlying the ability to solve each item. The resulting multifactorial models were tested against one-factor and two-factor models; using confirmatory factor analysis, the authors provided evidence for a five-factor structure of early numeracy at the end of kindergarten discerning the factors *patterning, seriation, counting, non-symbolic comparison*, and *symbolic number knowledge*. Additionally, these factors turned out to reliably predict later math performance in a curricular test in grade six. In particular, the authors found a unique association between *non-symbolic comparison, seriation, counting*, and *symbolic number knowledge*, and later mathematical skills, but no unique association for *patterning*.

Hirsch et al. (2018) also discussed that the basic numerical competences underlying early numeracy may depend on the content and range of topics addressed in the respective (large-scale assessment) tests. Consequently, studies considering other data sets based on different tests proposed different models to represent the multifactorial structure of early numeracy. In particular, Purpura and Lonigan (2013) considered data from a preschool assessment that addressed several basic numerical (e.g., counting forward/backward, symbolic and non-symbolic magnitude comparisons, etc.) but not geometric abilities. They found evidence for three highly correlated, yet distinct factors of early numeracy which they termed *numbering* (e.g., counting procedures, subitizing, and estimation), *relational* (e.g., ordinality, number comparison), and (arithmetic) *operational* (e.g., basic addition/subtraction problems) abilities reflecting the structure of early numeracy (see also Purpura and Lonigan, 2015).

This three-factor model was recently replicated and expanded based on data from an assessment of early numeracy that covered a broader range of tasks. In particular, Milburn et al. (2019) observed a four-factor model consisting of the factors *measurement*, *geometry*, *patterning*, and *numeracy*—with the latter conceptualized as a second-order factor that was further differentiated into *numbering*, *relations*, and *operations* as proposed by Purpura and Lonigan (2013). Another four-factor model of early numeracy in kindergarten children was reported by Hellstrand et al. (2020) who differentiated the factors of *symbolic and non-symbolic number knowledge*, *understanding mathematical relations*, *counting*, and *basic arithmetic* (see also Aunio and Räsänen, 2016). Thus, universal characteristics of early numeracy may be most likely derived by integrating results across different tests and different samples.

Longitudinal Stability of Children's Numerical Competences

Related to the question of the structure of early numeracy, findings regarding the stability of early numeracy or—more specifically—basic numerical competences underlying early numeracy in children's development are also limited. In fact—and in part due to the rapid development of numerical competences in early childhood—there is relatively little work in which the same measures were obtained repeatedly in a longitudinal design.

It is clear from the literature that overall (early) numeracy seems highly stable throughout early childhood and beyond (e.g., Bailey et al., 2014; Schmitt et al., 2017). However, in these studies, (early) numeracy was usually assessed by broad (standardized) math tests yielding a single-parameter score that subsumed performance on different numerical tasks. This may be problematic considering the multi-dimensional structure of (early) numeracy (e.g., Hirsch et al., 2018). As such, additional research is needed to better understand the stability of specific basic numerical competences underlying early numeracy and their structure over time. Previous longitudinal research on basic numerical competences has often been limited to very specific competences (e.g., non-symbolic magnitude comparison reflecting the approximate number system, ANS, e.g., Purpura and Simms, 2018) or measures that consist of very few items to describe a broad competence (e.g., counting and cardinality, e.g., van Marle et al., 2014; Purpura et al., 2017). However, the stability of the structure of basic numerical competences underlying early numeracy has been rarely evaluated so far. Even though Hellstrand et al. (2020) as well as Purpura and Lonigan (2013) were able to replicate their models in different age groups (e.g., in younger and older preschool children, Purpura and Lonigan, 2013), longitudinal (i.e., within-person) stability of a multifactorial structure of early numeracy has not yet been evaluated so far and thus remains unclear.

Generally, broader measures of numeracy of pre- and primary school children seem to be rather stable over time, as observed over periods of several months (e.g., Libertus et al., 2013; Chu et al., 2016; Nuutila et al., 2018) or even years (e.g., Aunola et al., 2004; Bailey et al., 2014; Schmitt et al., 2017). For instance, Aunola et al. (2004) obtained numeracy using a curriculum-based test six times over a period of 3 years and found scores to be highly interrelated. Moreover, Bailey et al. (2014) demonstrated that numeracy as assessed by standardized tests was highly stable across both short- (from first to fourth grade of primary school) and long-term periods (from first grade up to the age of 15 years). Using a state-trait model, the authors observed that a high degree of variance in numerical development over time is attributable to trait—rather than state—characteristics. Similarly, Jordan et al. (2007) measured children's number sense (a composite score of counting, number knowledge, estimation abilities, etc.) at several time points between kindergarten and 1st grade and found it to increase slightly but constantly (see also Jordan et al., 2006). Moreover, the predictive power of number sense measures for later mathematical achievement seemed to be stable throughout primary school (Jordan et al., 2010).

Additionally, previous studies also indicate that the stability of very specific numerical competences is high even in preschool children. For non-symbolic magnitude comparison (as a measure of ANS) some studies reported remarkably high test-retest correlations (e.g., Libertus et al., 2013; Toll et al., 2015; Chu et al., 2016; Purpura and Simms, 2018). For instance, Purpura and Simms (2018) measured ANS twice within 6 months of preschool and observed rather high stability, similar to results for symbolic and non-symbolic magnitude comparison and arithmetic abilities in primary school children (e.g., Göbel et al., 2014).

The Present Study

In sum, previous research has highlighted that early numeracy is multifactorial in that it is constituted by distinct basic numerical competences. However, studies explicitly investigating the structure of early numeracy are scarce. At the same time, evidence on the longitudinal (i.e., within-person) stability of basic numerical competences making up early numeracy mainly stems from studies which obtained either particularly broad and general or very specific measures of basic numerical competences so far. However, we are not aware of any study evaluating the longitudinal stability of a specific structure of basic numerical competences reflecting early numeracy within the same sample of children over time.

This is of particular interest because in early years of education it is likely that children's numerical development is dynamic as the numerical concepts become gradually more complex. As such, those basic numerical competences which constitute early numeracy might change. On the other hand, it is also well possible that early numeracy is a rather stable construct as its components reflect very basic building blocks of children's numerical competences reflected in curricular content strands (i.e., number sense and operations; measurement; geometry; data analysis and statistics; and patterning). Therefore, the aims of the present study were to investigate (i) the structure of early numeracy and (ii) to evaluate the stability of this structure over time in preschool children.

To pursue these aims, we relied on data from the Family Life Project (FLP), a large population-based prospective longitudinal study of children and families in predominantly low-income, non-urban communities in the United States. In addition to numerous aspects of child and family functioning, the FLP dataset contains data from the standardized math assessment developed for the Early Childhood Longitudinal Study-Kindergarten Cohort of 1998 (ECLS-K). Given that prior research and practice in math education indicated the need for a detailed *a priori* differentiation of numerical-mathematical competences (e.g., Aunio et al., 2006), we considered a confirmatory approach as particularly valuable to analyze the structure of early numeracy. Moreover, we tested whether this structure is stable across time by evaluating its validity to account for children's performance on the same test 1 year later.

MATERIALS AND METHODS

Participants

The original sample that made up the Family Life Project from which data are drawn was recruited when children were 2 months of age and comprised N = 1,292 children recruited to be representative of two of the four major geographical areas of high child rural poverty in the United States. Complex sampling procedures were used to recruit representative samples of non-urban areas of Pennsylvania and North Carolina, with intentional over sampling of low-income families and families of African American ethnicity. Five years later at PreK, over 70% of children (n = 911) participated in assessment, and in kindergarten over 80% of children (n = 1056) participated. We anticipate some of the difference between the two testing sessions was that not all children were enrolled in center-based care for PreK and were therefore more difficult to access whereas in kindergarten almost all children were enrolled in a school. Seventy percent of families had an average income of less than 200% of the poverty line. Additionally, 40% of mothers had a high school education (12 years of schooling) or less, while only 16% had at least 4 years of postsecondary education. A little more than half of the sample is White (57%) with the remainder of African American descent. Further details are available elsewhere (Vernon-Feagans et al., 2013).

Procedures

When children were approximately 60 months of age (M = 60.16 months, SD = 3.29) children were visited at their preschools by a trained data collector (or at home if they were not enrolled in center- or school-based care) to obtain the measure of early numeracy. In the spring of the child's kindergarten year (M_{age} = 71.40 months, SD = 3.36), they were again visited at their school by a trained data collector to re-assess early numeracy using the same test.

Measure of Math Ability

The ECLS-K math assessment was used to test children's early numeracy. The ECLS-K assessment uses a routing system to minimize administration time and most accurately assess their ability, and is a reliable and valid measure whose psychometric properties have been described elsewhere (Rock and Pollack, 2002). The same assessment was used in both pre-kindergarten (PreK) and kindergarten. All participants receive a series of 14 routing items. If participants scored 8 or lower on routing items, they are directed to the "low" block; if higher than 8, an additional 4 routing items were administered. If participants correctly respond to between 9 and 11 items, they were routed to the "medium" block; if 12 or higher, they were routed to the "high" block. The low block had 18 items plus the 14 routing items, the medium block had 25 items plus the 18 routing items, and the high block had 32 items plus the 18 routing items. Children were routed to high, medium, or low blocks on the basis of the number of items they got correct on the routing section of the assessment (all other things being equal). In particular, there was no adjustment for any child characteristics (e.g., age and sex) such that any child had equal probability of being routed to any block. In PreK, 93.2% of participants (N = 849) were routed to the "low" block, 5.5% of participants (N = 50) were routed to the "medium" block, and 1.3% of participants (N = 12) were routed to the "high" block. In kindergarten, 51.8% of participants (N = 537) were routed to the "low" block in PreK, 30.2% of participants (N = 319) were routed to the "medium" block, and 18.9% of participants (N = 200) were routed to the "high" block.

Analytic Strategy

We conducted confirmatory factor analysis (CFA) to analyze the multifactorial structure of basic numerical competences underlying early numeracy. In particular, we specified and evaluated a one-factor model representing early numeracy as a unitary construct and compared it to a multifactorial (six-factor) model in which items from the ECLS-K were classified based on the basic numerical competences necessary to solve each item. The categories for item coding (six basic numerical competences) were derived from the psychometric report of the ECLS-K (Rock and Pollack, 2002). This classification is mainly based on curriculum standards and reflects the way in which the ECLS-K was designed. Therefore, it takes the specific characteristics of this test into account while it also shows structural and conceptual similarities to previous multifactorial models of early numeracy. Moreover, it covers distinct basic numerical competences that have already been investigated in previous early numeracy research (for an overview see Table 1 in Hirsch et al., 2018). To evaluate how well the data fit the theorized models, we considered the cutoff criteria presented by Hu and Bentler (1999): A well-fitting model was expected to have a Comparative Fit Index (CFI) $>$ 0.95, and Root Mean Squared Error of Approximation (RMSEA) $<$ 0.08. Models were estimated in Mplus (Muthén and Muthén, 2017) and used the Weighted Least Squared Means and Variances (WLSMV) estimator. Prior research suggested WLSMV is appropriate for ordinal variables, and is less biased than are other estimators (Li, 2016). In a second step, measurement invariance was tested to establish whether the same constructs could be established 1 year later in kindergarten to evaluate the stability of early numeracy. Adequate model fit was determined by use of a chi-square difference test and whether CFI changed more than 0.002 (Meade et al., 2008).

Participants were included in analyses if they took part in the PreK wave of data collection and were routed to the "low" block (N = 849; 93.2%). Missing data at the re-test in Kindergarten was accounted for using Full Information Maximum Likelihood estimation. This approach takes into account the covariance matrix for all available data on the independent variables to estimate parameters and standard errors and provides more

TABLE 1 | Descriptive analyses for items from ECLS-K math assessment.

		Pre-K				Kindergarten			
Item	**Item type**	***N***	**% correct**	***SD***	**Factor loading**	***N***	**% correct**	***SD***	**Factor loading**
17	Arithmetic	846	0.31	0.46	0.57	830	0.51	0.50	0.85
26	Arithmetic	846	0.42	0.49	0.51	440	0.65	0.48	0.36
11	Arithmetic	846	0.16	0.37	0.57	440	0.49	0.50	0.34
25	Arithmetic	846	0.18	0.39	0.47	440	0.20	0.40	0.29
31	Data Analysis/statistics	846	0.57	0.50	0.92	440	0.83	0.37	0.90
30	Data Analysis/statistics	846	0.48	0.50	0.85	440	0.75	0.43	0.93
18	Number sense	846	0.76	0.43	0.89	440	0.96	0.20	0.63
19	Number sense	846	0.63	0.48	0.90	440	0.91	0.29	0.56
5	Number sense	846	0.43	0.50	0.80	830	0.87	0.33	0.76
20	Number sense	846	0.20	0.40	0.80	440	0.51	0.50	0.45
4	Number sense	846	0.28	0.45	0.68	830	0.76	0.43	0.54
6	Number sense	846	0.11	0.32	0.60	830	0.61	0.49	0.78
16	Number sense	846	0.86	0.35	0.63	440	0.96	0.19	0.36
7	Number sense	846	0.11	0.31	0.56	830	0.61	0.49	0.85
15	Number sense	846	0.94	0.23	0.61	440	0.98	0.14	0.27
9	Number sense	846	0.14	0.35	0.47	830	0.62	0.49	0.75
32	Number sense	846	0.37	0.48	0.25	440	0.40	0.49	0.17
29	Patterning	846	0.49	0.50	0.57	440	0.69	0.46	0.32
2	Patterning	846	0.73	0.45	0.51	830	0.91	0.29	0.59
8	Patterning	846	0.58	0.49	0.57	830	0.82	0.39	0.64
1	Patterning	846	0.53	0.50	0.47	830	0.76	0.43	0.61
22	Patterning	846	0.44	0.50	0.41	440	0.61	0.49	0.36
3	Patterning	846	0.41	0.49	0.30	830	0.65	0.48	0.56
24	Patterning	846	0.38	0.48	0.30	440	0.57	0.50	0.24
10	Excluded	846	0.01	0.12	N/A	830	0.17	0.38	N/A
12	Excluded	846	0.15	0.36	N/A	830	0.31	0.46	N/A
13	Excluded	846	0.05	0.23	N/A	830	0.22	0.41	N/A
14	Excluded	846	0.02	0.14	N/A	830	0.14	0.34	N/A
21	Excluded	846	0.90	0.30	N/A	440	0.96	0.20	N/A
23	Excluded	846	0.33	0.47	N/A	440	0.52	0.50	N/A
27	Excluded	846	0.04	0.20	N/A	440	0.07	0.25	N/A
28	Excluded	846	0.11	0.32	N/A	440	0.18	0.39	N/A

accurate estimates of regression coefficients than do listwise deletion or mean replacement (Enders, 2011).

RESULTS

Descriptive Statistics

Descriptive statistics for items from the ECLS-K math assessment are shown in **Table 1**. Correlations between all variables are shown in **Table 2**.

Confirmatory Factor Analysis of Pre-kindergarten Basic Numerical Competences

In line with theoretical considerations and the psychometric report from the ECLS-K assessment items were coded as assessing one of six basic numerical competences: Patterning (6 items), Number Sense (12 items), Arithmetic (7 items), Geometry (3 items), Measurement (2 items), and Data Analysis/Statistics (2 items). **Table 3** provides a more detailed description of the items. In our first model, we were interested to test whether there was sufficient distinction of separate constructs to justify the operationalization of six separate numerical competencies. To do so, we first tested whether all items loaded onto a single factor. The resulting model did not fit the data well, χ^2 = 1021.04, $p < 0.001$; RMSEA = 0.038, 90% CI [0.035, 0.041]; CFI = 0.904. We then tested a second model in which the data were fit to the six purported constructs. The resulting model also did not fit the data at our desired levels, χ^2 = 761.62, $p < 0.001$; RMSEA = 0.029, 90% CI [0.025, 0.032]; CFI = 0.946; however, the model fit was improved significantly over the single-factor model, χ^2 = 261.61, $p < 0.001$. The model was further specified: Items that did not load onto any of the constructs at a level of $p = 0.10$ or lower, as well as items without sufficient variance in responses (i.e., items which were correctly/incorrectly solved by almost all children) were dropped (8 items in total). Additionally, modification indices suggested items in the *measurement* (1 item) and

TABLE 2 | Correlations among latent variables.

		1	2	3	4	5	6	7	8
1	Patterning PreK	–							
2	Number Sense PreK	0.70***	–						
3	Arithmetic PreK	0.89***	0.66***	–					
4	Data Analysis/Statistics PreK	0.65***	0.57***	0.71***	–				
5	Patterning K	0.86***	0.57***	0.70***	0.58***	–			
6	Number Sense K	0.69***	0.77***	0.63***	0.53***	0.71***	–		
7	Arithmetic K	0.71***	0.54***	0.84***	0.57***	0.82***	0.74***	–	
8	Data Analysis/Statistics K	0.13	0.18**	0.32**	0.25***	0.42***	0.39***	0.48***	–

****$p < 0.001$, **$p < 0.01$.*

geometry (2 items) categories loaded onto the *patterning* category, and one item from the *number sense* category loaded onto the *arithmetic* category. These modifications seemed theoretically justified. In particular, structural similarities were found in the geometry and patterning items as both required visuospatial recognition of shapes and patterns and the measurement item was presented in a similar way as patterning and geometry items (i.e., children had to select the correct solution from a set of four alternatives). The number sense item was conceptualized as assessing counting, but it required addition of three sets of objects which indeed seemed related to arithmetic. We therefore decided to re-specify the model accordingly. The resulting 4-factor model fit the data well, $\chi^2 = 474.26$, $p < 0.001$; RMSEA = 0.033, 90% CI [0.029, 0.038], CFI = 0.962. Because the updated model had items missing, model fit could not be formally compared to the 6-factor model using a log-likelihood difference test because models were no longer nested; however, model fit for the 4-factor model was adequate according to conventional norms whereas it was not for the 6-factor model. Fit indices for all models are provided in **Table 4**.

The 4-factor model included factors for *patterning and geometry*, *number sense*, *arithmetic*, as well as *data analysis and statistics* (see **Figure 1** for illustrating example items, see **Table 1** for item properties). Factor loadings for each factor are presented in **Table 1**. The variances of latent variables were significant (patterning, $\sigma^2 = 0.23$, $p < 0.001$; number sense, $\sigma^2 = 0.43$, $p < 0.001$; arithmetic, $\sigma^2 = 0.26$, $p < 0.001$; data analysis/statistics, $\sigma^2 = 0.71$, $p < 0.001$). The four latent variables were correlated with one another, and correlations among latent variables are presented in **Table 2**. Children showed the highest levels of understanding *patterning and geometry* (M = 53.82% correct) and *data analysis/statistics* (M = 52.60% correct), then *number sense* (M = 46.76% correct), and *finally arithmetic* (M = 29.19% correct).

Stability of Basic Numerical Competences Underlying Early Numeracy

To test the validity and stability of the numerical competences established in PreK, a series of models were run to test longitudinal measurement invariance of numerical competences in kindergarten. We first tested configural invariance to examine whether the items that represented the identified constructs in PreK continued to do so in kindergarten. A confirmatory model in which the same four factors (i.e., patterning/geometry, number sense, arithmetic, and data analysis/statistics) were simultaneously estimated in PreK and kindergarten. The model fit the data well, $\chi^2 = 1531.184$, $p < 0.0001$; RMSEA = 0.023, 90% CI [0.021, 0.026], CFI = 0.950, such that configural invariance could be concluded indicating that the same items represented the identified constructs in PreK and kindergarten.

Metric invariance was then tested to examine the relative contribution of the items within factors, in that the coefficients of items in each factor were set to be equal across administration (that is, in PreK and kindergarten). Model fit was acceptable, $\chi^2 = 1935.951$, $p < 0.0001$; RMSEA = 0.031, 90% CI [0.029, 0.033], CFI = 0.910; however, the chi-square test of model difference was significant ($\chi^2 = 204.604$, $p < 0.001$) and CFI changed markedly more than 0.002, indicating metric invariance was not held.

DISCUSSION

Prior empirical work has suggested that early numeracy might be better represented as a multidimensional construct made up of distinct basic numerical competences than a single unitary construct (e.g., Dowker, 2008). However, multidimensional conceptualizations are rare. Additionally, evidence on the longitudinal (i.e., within-person) stability of a specific structure of basic numerical abilities underlying early numeracy over time is limited. The present study aimed at complementing prior research by evaluating the longitudinal stability of the structure of basic numerical competences in a large longitudinal data set of young children in the United States. Primary aims were (i) to test the structure of basic numerical competences constituting early numeracy in 5-year-old children and (ii) to evaluate the stability of this structure over 1 year from PreK through kindergarten. In the following, we will discuss these aspects in turn.

Structure of Basic Numerical Competences Underlying Early Numeracy

Early numeracy has been typically considered a unitary or two-factorial construct in previous studies (e.g., Aunio et al., 2004; Jordan et al., 2010). However, there is evidence also using

TABLE 3 | Item descriptions.

Item	Item type	Item description
11	Arithmetic	(Object-based) addition[a]
17	Arithmetic	(Object-based) addition
25	Arithmetic	(Object-based) addition
26	Arithmetic	(Object-based) subtraction
30	Data Analysis/statistics	Graph reading[a]
31	Data Analysis/statistics	Graph reading
4	Number sense	Counting forward
5	Number sense	Identify a number (symbolic number knowledge)
6	Number sense	Identify a number (symbolic number knowledge)
7	Number sense	Identify the *n*th object (ordinality)
9	Number sense	Complete a number series (seriation/number order)
15	Number sense	Count a set
16	Number sense	Numeral recognition[a]
18	Number sense	Identify a written number (symbolic number knowledge)
19	Number sense	Identify a written number (symbolic number knowledge)
20	Number sense	Identify a written number (symbolic number knowledge)
32	Number sense	Estimation (non-symbolic)
1	Patterning	Match a pattern of objects/shapes to a different pattern from a set[a]
2	Patterning	Match a pattern of objects/shapes to a different pattern from a set
3	Patterning	Match a pattern of objects/shapes to a different pattern from a set
8	Patterning	Choose a shorter/larger object from a set (geometry/length concept)
22	Patterning	Match a pattern of objects/shapes to a different pattern from a set
24	Patterning	Match a geometric shape to a different shape from a set (geometry)
29	Patterning	Measure length of an object (measurement)
12	Excluded (arithmetic)	(Object-based) subtraction
13	Excluded (arithmetic)	(Object-based) addition
27	Excluded (arithmetic)	(Object-based) addition
28	Excluded (arithmetic)	(Object-based) subtraction
21	Excluded (geometry)	Recognize a geometric shape
14	Excluded (measurement)	Operate with money (measurement)
10	Excluded (number sense)	Complete a number series (seriation/number order)
23	Excluded (patterning)	Match a pattern of objects/shapes to a different pattern from a set

[a] *see* ***Figure 1*** *for a generic example.*

large-scale assessment data that early numeracy in preschool years may be constituted by more than two basic numerical competences (Purpura and Lonigan, 2013; Hirsch et al., 2018; Milburn et al., 2019; Hellstrand et al., 2020). In these studies, items from large-scale assessments of early numeracy were used to specify and evaluate multifactorial models by means of confirmatory factor analysis. However, these models differed in content and structure. In particular, Purpura and Lonigan (2013) established a three-factor model of early numeracy with factors for *numbering*, *relation*, and *operation* competences. Milburn et al. (2019) extended this model by adding another three distinct factors for *measurement*, *geometry*, and *patterning* competences. Another four-factor model with *symbolic/non-symbolic number knowledge*, *numerical relations*, *basic arithmetic*, and *counting* competences was recently presented by Hellstrand et al. (2020), and Hirsch et al. (2018) substantiated a five-factor model reflecting *patterning*, *seriation*, *non-symbolic comparison*, *counting*, and *symbolic number knowledge* abilities. In contrast, the present study using the ECLS-K math assessment identified early numeracy as constituted by four basic numerical competences reflected in a confirmatory factor analytic approach. The one- and six-factor confirmatory models evaluated did not yield adequate model fit; rather, our results provided evidence for four basic numerical competences underlying early numeracy as assessed by the ECLS-K math assessment: (i) *patterning and geometry*, (ii) *number sense*, (iii) *arithmetic*, as well as (iv) *data analysis and statistics*.

TABLE 4 | Model fit information.

Model	χ^2	df	CFI	RMSEA [90%CI]
One factor	1021.04	464	0.904	0.038 [0.035, 0.041]
Six factor	761.62	449	0.946	0.029 [0.025, 0.032]
Four factor	474.3	246	0.962	0.033 [0.029, 0.038]

Comparing our results with those of previous studies revealed that the basic numerical competences specified in the different models were quite similar indeed. Most obviously, we also found a factor for basic *arithmetic operations* similar to Purpura and Lonigan (2013), Milburn et al. (2019), and

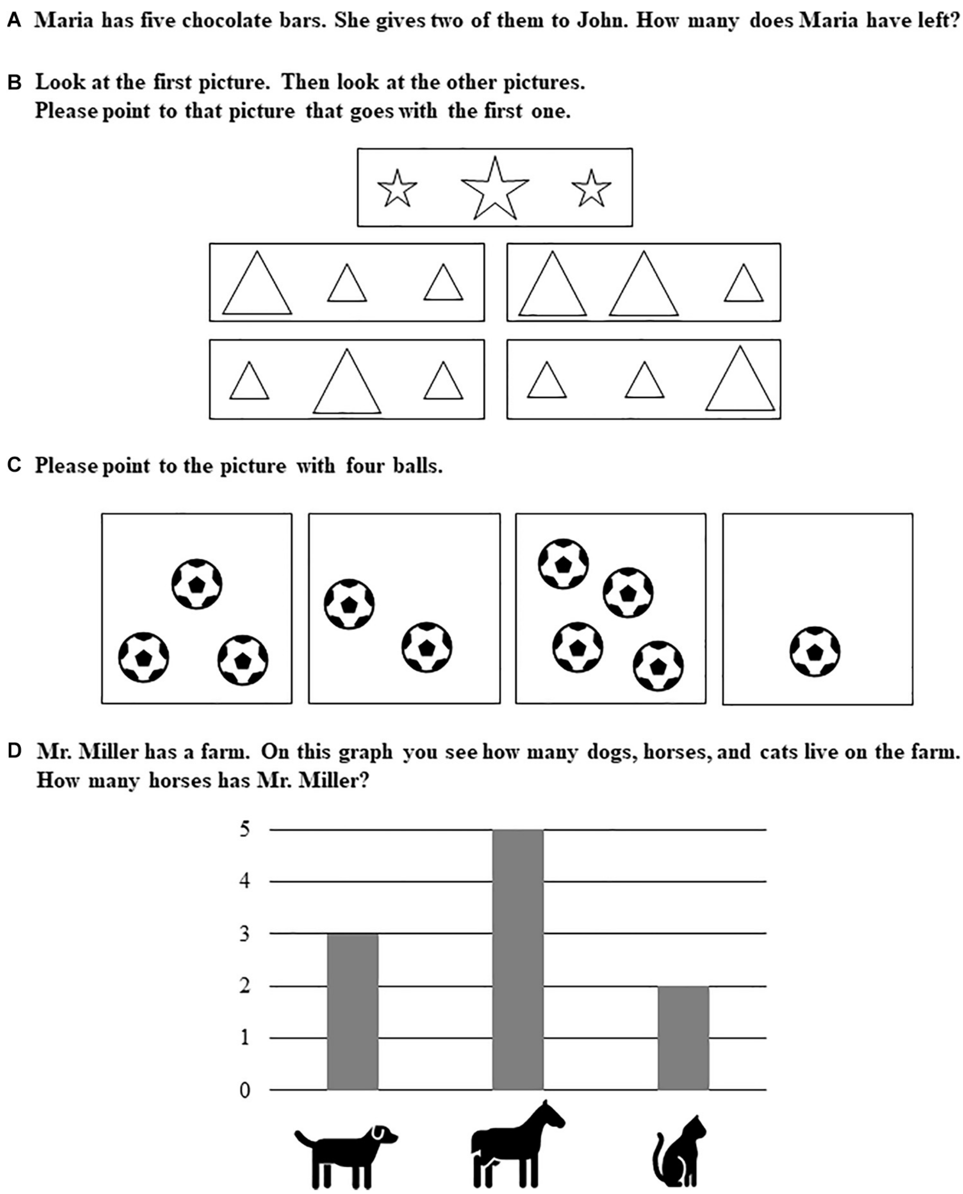


FIGURE 1 | Generic example items for **(A)** arithmetic, **(B)** patterning/geometry, **(C)** number sense, and **(D)** data analysis/statistics. Instructions were read out by the investigators while the items were shown to the children on a separate sheet in an open-bound spiral notebook.

Hellstrand et al. (2020), and a factor for *patterning* (which here also included *geometry* and *measurement*) as did Hirsch et al. (2018) and Milburn et al. (2019). Although several factors of models reported in other studies did not directly correspond to those observed in the present study, a more detailed comparison of item contents from this and previous studies revealed (**Table 3**) that they seem to be in part subsumed in our *number sense* factor. In particular, *number sense* was mainly assessed by symbolic number knowledge and counting items and therefore largely overlaps with the *numbering* factor in Purpura and Lonigan (2013) and Milburn et al. (2019). Additionally, it comprised a few items on seriation, ordinality, and estimation which overlaps with content of the *relations* factor in the models of Purpura and Lonigan (2013) and Milburn et al. (2019), or the *relations* and *symbolic/non-symbolic number knowledge* factors in the model of Hellstrand et al. (2020). Furthermore, our *number sense* factor may reflect a conjunction of four factors of the model by Hirsch et al. (2018), namely *counting*, *seriation*, *symbolic number knowledge*, and *non-symbolic comparison*.

Despite these significant commonalities, our model of early numeracy differed in at least three notable aspects from previously identified multifactorial models. First, as already indicated above, *number sense* described a rather broad factor compared to more specific numerical competences the other studies specified. However, this may be due to the fact that the ECLS-K math assessment was explicitly designed to measure number sense broadly, which made it difficult to identify more specific competences based on the number sense items as too few items were available. For instance, we might have further specified a specific factor for seriation, but only one item actually addressed seriation in the ECLS-K assessment. As such, it was more appropriate to summarize such (single) items under a more general *number sense* factor.

Second, a similar reason may also explain why we did not find distinct factors for *patterning*, *geometry*, and *measurement* as did Milburn et al. (2019). In particular, their model comprised four patterning, seven geometry, and six measurement items, while we identified only five patterning, two geometry and one measurement items. Accordingly, limited variance on *geometry* and *measurement* in our data may have been best explained by *patterning*. Additionally, geometry and measurement items were structurally very similar to the patterning items.

Lastly, our model suggests that *data analysis and statistics* seemed to represent another distinct basic numerical competence that may already emerge in preschool years. This competence describes children's ability to read and draw inferences of graphical representations of data. To the best of our knowledge, *data analysis and statistics* has not yet been reported in other multifactorial models of preschoolers' early numeracy.

At the same time, however, all differences between models discussed here may not be unexpected as different assessments of early numeracy with (partially) different mathematical-numerical content assessed may lead to the identification of different basic numerical competences that constitute the construct of early numeracy (cf. Hirsch et al., 2018). This may be particularly so given that the large-scale assessment tests considered in some of the studies, including this one, originally intended to reflect a broad assessment of early numeracy as it was defined in curricular standards of different educational systems (e.g., the Netherlands vs. United States). However, they were not designed to explicitly measure a universal structure of specific basic numerical competences. Nevertheless, we think that it is these comparisons across studies on different samples and different assessments that offer a promising way to gain a comprehensive multidimensional view on early numeracy.

Stability of Basic Numerical Competences Constituting Early Numeracy

After substantiating the four-factor structure of early numeracy assessed in the ECLS-K math assessment, we evaluated the stability of this structure by testing measurement invariance of the factor structure with a subsample of children that were assessed twice on the same test from PreK to kindergarten. Our analyses revealed configural but not metric invariance, indicating that we were able to identify the same four factors (with the same items) of early numeracy in kindergarten as in PreK, but within that year the relative contributions of items to the factors (i.e., factor loadings) changed. In other words, when children became older some items became stronger (or weaker) indicators of the respective basic numerical competences. Most likely, this reflects that children became more proficient in math and were better able to solve the respective numerical tasks in kindergarten than in PreK. This is also reflected in the smaller number of children routed to the low block in the ECLS-K math assessment. Importantly, however, the four basic numerical competences constituting early numeracy in PreK remained stable to kindergarten with all factor loadings of indicators on a significant and meaningful level. Taken together, these findings suggest that a structure of early numeracy that consists of four correlated factors (i.e., *patterning and geometry, number sense, arithmetic*, as well as *data analysis and statistics*) continues to be refined and improved over time.

However, we cannot conclude whether the factor structure we established here remains stable beyond the preschool years. In particular, the curricula to which children were exposed in PreK and kindergarten were likely more comparable than those of kindergarten and first or second grade; indeed, prior investigations have suggested that many kindergarten teachers spend the majority of their time teaching students what they already learned in preschool (Engel et al., 2013). As such, it is possible that after the beginning of formal schooling when numerical/mathematical content becomes increasingly complex and math instruction more formal, children alter and restructure their early numeracy more substantially. The fact that we did not observe metric invariance of the evidenced four-factor structure may already indicate substantial changes to take place. However, some prior evidence suggests stability of early numeracy at least through the early years of education (e.g., Hellstrand et al., 2020, established their four-factor model in samples of kindergarten, first, and second grade children). Nevertheless, it should be subject to future research to investigate a multifactorial structure of basic numerical competences and follow its development longitudinally across a longer period of time than it was done in the present study. As such, the four basic numerical competences we established in the present study seem to be an essential foundation of children's early numeracy over the course of preschool. This may have practical and theoretical implications.

Research on intelligence has shown that analyzing the contributions of specific cognitive abilities to school achievement is more informative than considering only a general *g*-factor of intelligence (e.g., Gustafsson and Balke, 1993; Calvin et al., 2010). Similar to these results, the present study highlights the need and value of a more differentiated view on early numeracy in children (cf. Cowan and Powell, 2014). Information on specific basic numerical competences that make up children's early numeracy as well as their stability during preschool is not only essential for research on numerical cognition, but also in broader educational contexts. For instance, when it comes to diagnose children with mathematical learning difficulties it is important to identify their problems and deficits as early and as specific as possible to initiate targeted interventions (e.g., Geary et al., 2009). This study provides empirical evidence that may help to improve both the diagnostic process itself but also the development of subsequent interventions as it allows for the specification of basic numerical competences making up early numeracy. Based on this, it should be possible to develop diagnostic tools to specifically assess and intervene upon these basic numerical competences. As the present study is among the first of its kind, it must be acknowledged that implications for education are tentative, and further research is required to substantiate both the generalizability and longitudinal relevance of the present findings.

Limitations and Perspectives

The present study was inspired by the multifactorial model of early numeracy proposed by Hirsch et al. (2018). So far, previous research identified several basic numerical competences often using different tests and sometimes different labels but more or less corresponding to each other when considering underlying basic numerical competences (for an overview see Table 1 in Hirsch et al., 2018). As such, it would be desirable to develop a consensual conceptualization of early numeracy which serves as a framework in future research on basic numerical competences and their long-term relevance.

In the present study, early numeracy was found to be constituted by the four basic numerical competences *patterning and geometry, number sense, arithmetic*, as well as *data analysis and statistics*. As such, we propose that patterning and geometry are also important domains of early numeracy. It should be noted, however, that this goes beyond previous studies that primarily focused on number-related and operational content (e.g., counting, cardinality understanding, addition/subtraction, etc.) when conceptualizing early numeracy (e.g., Purpura et al., 2011; Purpura and Lonigan, 2013; Nelson and McMaster, 2018). Nevertheless, consideration of patterning and/or geometry as important to early numeracy is in line with other studies (e.g., Jordan et al., 2006; Clements and Sarama, 2011; Polignano and Hojnoski, 2012; Rittle-Johnson et al., 2015; Hirsch et al., 2018; Milburn et al., 2019). Moreover, it is also in line with curricular strands on early math education which typically incorporate patterning and early geometry (National Council of Teachers of Mathematics, 1989, 2000).

Moreover, there are further limitations to be considered when interpreting the current results, which may – at the same time – provide interesting avenues for future research. First, we did neither evaluate within-sample effects (e.g., child gender) as we expect the stability of constructs should not vary as a result of demographic characteristics, nor did we consider influences of other domain-general abilities (e.g., language ability and executive functions) on the development of basic numerical competences in our model. Prior research suggested significant interrelations of these variables and several (general or specific) measures of math ability (e.g., Blair et al., 2008; Praet et al., 2013). However, one might expect that they influence specific basic numerical competences differentially. For instance, Purpura et al. (2017) investigated how different components of executive functions (i.e., response inhibition/inhibitory control, cognitive flexibility, and working memory) predicted performance on various tasks on early numeracy. Response inhibition and cognitive flexibility turned out to predict, among others, measures that would correspond to the *number sense* factor in our model (e.g., subitizing, counting, number ordering, and cardinality). In contrast, working memory primarily predicted performance in tasks that required to execute multiple steps or keeping track of intermediate results (e.g., computations). As such, it may mostly be related to the *arithmetic* operations factor of the present model. Similarly, language ability may be most strongly related to competences, which were assessed using word problems or other largely text-based items, that is, *arithmetic* and *data analysis/statistics* in the current model. Nevertheless, we can only speculate about these potential influences so far and it should be investigated in future studies to which degree variance in specific basic numerical competences may be actually explained by domain-general variables. At the same time, however, it should also be noted that using large-scale assessment data for the purpose of secondary data analysis may be often constrained insofar as further potentially interesting variables (e.g., covariates, further indices of achievement) were not addressed during data collection.

Second, the results of the present study might not be generalizable to a wider population. While we leverage quite a large sample, findings may primarily apply to a certain population due to the sample characteristics (i.e., children from low-income families living mainly in non-urban regions in the United States). In particular, we were able to investigate the stability of the factor structure of early numeracy only in those children who were routed to the low-ability block of tasks in the math assessments in PreK and kindergarten as this was the largest group in the longitudinal sample. As such, early numeracy might be less stable in children with higher abilities or steeper learning curves in math. It is conceivable, for instance, that those children develop further numerical competences during preschool due to a differentiation of their number sense abilities (e.g., into symbolic and non-symbolic numerical abilities). This process might be delayed in the low-performing children we considered in the present study.

Finally, as noted earlier, future research may continue to investigate the multifactorial structure of early numeracy longitudinally across longer time periods. This would further allow to evaluate the longitudinal relevance of specific early numerical and later mathematical competences. In particular, prior research primarily specified the predictive power of early basic numerical competences for later general math achievement. For instance, early symbolic numerical competences (e.g., symbolic number knowledge) were shown to predict later math achievement (see e.g., Schneider et al., 2017 for a review). Additionally, Hirsch et al. (2018) found all but one of the competences specified in their 5-factor model of early numeracy to predict math achievement in grade 6. However, it is currently not clear how basic numerical competences establishing a multifactorial structure of early numeracy may relate to a differentiated multifactorial structure of basic numerical competences and/or more advanced mathematical competences (e.g., fraction understanding) established later within the same individuals (i.e., beyond the period of 1 year we covered in this study). Investigating this in more depth would provide more detailed knowledge on the long-term development of basic numerical competences and of potential variation in their interrelations as well as their relation to more advanced mathematical abilities.

Moreover, most recent studies in basic numerical cognition research face the issue of rather small sample sizes. In particular, Kolkman et al. (2013) argued the need for replications of findings using large-scale data. In this study, we specified the structure and stability of latent basic numerical competences underlying a broad curriculum-based assessment and discussed remarkable overlap with prior studies taking a similar approach. We thus see a specific advantage of confirmatory analyses considering large-scale assessment data of early numeracy in general and when it comes to evaluate basic numerical competences with different tests in different samples in particular.

CONCLUSION

To the best of our knowledge, this study is the first attempt to replicate a multifactorial model of early numeracy using large-scale assessment data across the PreK and kindergarten years. Importantly, we found further evidence that early numeracy in preschool children is constituted by different basic numerical competences. In particular, we found early numeracy to be reflected by the following four basic numerical competences: (i) *patterning and geometry*, (ii) *number sense*, (iii) *arithmetic*, as well as (iv) *data analysis and statistics*. Although labeled differently, we were able to replicate most factors proposed in prior studies on the multifactorial structure of early numeracy. Moreover, we provided first evidence for the stability of the structure of basic numerical competences constituting early numeracy from PreK to kindergarten. This highlights the role of early numeracy and its underlying basic numerical competences as an important foundation for later numerical-mathematical development. The present findings imply, that preschool education should recognize the multidimensional nature of early numeracy and specifically foster children's mastery of basic numerical competences.

FAMILY LIFE PROJECT KEY INVESTIGATORS

The Family Life Project Key Investigators include Lynne Vernon-Feagans, The University of North Carolina, Mark Greenberg, The Pennsylvania State University, Martha Cox, The University of North Carolina, Clancy Blair, New York University, Peg Burchinal, The University of North Carolina, Michael Willoughby, The University of North Carolina, Patricia Garrett-Peters, The University of North Carolina, Roger Mills-Koonce, The University of North Carolina.

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available as some restrictions will apply. Data are from the Family Life Project study whose authors may be contacted at https://flp.fpg.unc.edu/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Boards of The University of North Carolina and The Pennsylvania State University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

DB, AR, and KM conceptualized the study. DB and AR wrote the main manuscript text. AR performed the statistical analyses. CB and KM provided critical editorial feedback and thoughtful revision to the text. All authors reviewed and approved the manuscript as written.

FUNDING

Support for this research was provided by the National Institute of Child Health and Human Development grants R01 HD51502 and P01 HD39667 with co-funding from the National Institute on Drug Abuse. The publication was supported by Open Access Publishing Fund of University of Tübingen.

ACKNOWLEDGMENTS

The Family Life Project Key Investigators include Lynne Vernon-Feagans and Martha J. Cox, The University of North Carolina at Chapel Hill; Clancy Blair, New York University; Margaret R. Burchinal, The University of North Carolina at Chapel Hill; Linda Burton, Duke University; Keith Crnic, Arizona State University; Ann Crouter, Pennsylvania State University; Patricia Garrett-Peters, The University of North Carolina at Chapel Hill; Mark T. Greenberg, Jennifer L. Frank, Cynthia Stifter, Emily Werner, and Stephanie Lanza, Pennsylvania State University; W. Roger Mills-Koonce, The University of North Carolina at Greensboro; and Michael Willoughby, RTI International. We also thank the many families and research assistants who made this study possible.

REFERENCES

Anders, Y., Rossbach, H.-G., Weinert, S., Ebert, S., Kuger, S., Lehrl, S., et al. (2012). Home and preschool learning environments and their relations to the development of early numeracy skills. *Early Child. Res. Quart.* 27, 231–244. doi: 10.1016/j.ecresq.2011.08.003

Aunio, P., Ee, J., Lim, S. E. A., Hautamäki, J., and Van Luit, J. (2004). Young children's number sense in finland, hong kong and singapore. *Int. J. Early Years Educ.* 12, 195–216. doi: 10.1080/0966976042000268681

Aunio, P., Hautamäki, J., Heiskari, P., and Luit, J. E. H. V. (2006). The early numeracy test in finnish: children's norms. *Scand. J. Psychol.* 47, 369–378. doi: 10.1111/j.1467-9450.2006.00538.x

Aunio, P., and Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learn. Individ. Diff.* 20, 427–435. doi: 10.1016/j.lindif.2010.06.003

Aunio, P., and Räsänen, P. (2016). Core numerical skills for learning mathematics in children aged five to eight years – a working model for educators. *Eur. Early Childhood Educ. Res. J.* 24, 684–704. doi: 10.1080/1350293X.2014.996424

Aunola, K., Leskinen, E., Lerkkanen, M.-K., and Nurmi, J.-E. (2004). Developmental dynamics of math performance from preschool to grade 2. *J. Educ. Psychol.* 96, 699–713. doi: 10.1037/0022-0663.96.4.699

Bailey, D. H., Watts, T. W., Littlefield, A. K., and Geary, D. C. (2014). State and trait effects on individual differences in children's mathematical development. *Psychol. Sci.* 25, 2017–2026. doi: 10.1177/0956797614547539

Baker, D., Knipe, H., Collins, J., Leon, J., Cummings, E., Blair, C., et al. (2010). One hundred years of elementary school mathematics in the United States: a content analysis and cognitive assessment of textbooks from 1900 to 2000. *J. Res. Math. Educ.* 41, 383–423.

Blair, C., Knipe, H., and Gamson, D. (2008). Is there a role for executive functions in the development of mathematics ability? *Mind Brain Educ.* 2, 80–89. doi: 10.1111/j.1751-228X.2008.00036.x

Calvin, C. M., Fernandes, C., Smith, P., Visscher, P. M., and Deary, I. J. (2010). Sex, intelligence and educational achievement in a national cohort of over 175,000 11-year-old schoolchildren in England. *Intelligence* 38, 424–432. doi: 10.1016/j.intell.2010.04.005

Chu, F. W., van Marle, K., and Geary, D. C. (2016). Predicting children's reading and mathematics achievement from early quantitative knowledge and domain-general cognitive abilities. *Front. Psychol.* 7:775. doi: 10.3389/fpsyg.2016.00775

Cirino, P. T. (2011). The interrelationships of mathematical precursors in kindergarten. *J. Exp. Child Psychol.* 108, 713–733. doi: 10.1016/j.jecp.2010.11.004

Clements, D. H., and Sarama, J. (2011). Early childhood teacher education: the case of geometry. *J. Math. Teach. Educ.* 14, 133–148. doi: 10.1007/s10857-011-9173-0

Cowan, R., and Powell, D. (2014). 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

Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Dev. Sci.* 11, 650–654. doi: 10.1111/j.1467-7687.2008.00713.x

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

Enders, C. K. (2011). Analyzing longitudinal data with missing values. *Rehabil. Psychol.* 56, 267–288. doi: 10.1037/a0025579

Engel, M., Claessens, A., and Finch, M. A. (2013). Teaching students what they already know? The (mis)alignment between mathematics instructional content and student knowledge in kindergarten. *Educ. Eval. Policy Anal.* 35, 157–178. doi: 10.3102/0162373712461850

Geary, D. C., Bailey, D. H., Littlefield, A., Wood, P., Hoard, M. K., and Nugent, L. (2009). First-grade predictors of mathematical learning disability: a latent class trajectory analysis. *Cogn. Dev.* 24, 411–429. doi: 10.1016/j.cogdev.2009.10.001

Göbel, S. M., Watson, S. E., Lervåg, A., and Hulme, C. (2014). Children's arithmetic development: it is number knowledge, not the approximate number sense, that counts. *Psychol. Sci.* 25, 789–798. doi: 10.1177/0956797613516471

Gustafsson, J.-E., and Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multi. Behav. Res.* 28, 407–434. doi: 10.1207/s15327906mbr2804_2

Hellstrand, H., Korhonen, J., Räsänen, P., Linnanmäki, K., and Aunio, P. (2020). Reliability and validity evidence of the early numeracy test for identifying children at risk for mathematical learning difficulties. *Int. J. Educ. Res.* 102:101580. doi: 10.1016/j.ijer.2020.101580

Hirsch, S., Lambert, K., Coppens, K., and Moeller, K. (2018). Basic numerical competences in large-scale assessment data: structure and long-term relevance. *J. Exp. Child Psychol.* 167, 32–48. doi: 10.1016/j.jecp.2017.09.015

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equat. Model.* 6, 1–55. doi: 10.1080/10705519909540118

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., 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

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. 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.

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

Kroesbergen, E. H., Van Luit, J. E. H., Van Lieshout, E. C. D. M., Van Loosbroek, E., and Van de Rijt, B. A. M. (2009). Individual differences in early numeracy: the role of executive functions and subitizing. *J. Psychoeduc. Assess.* 27, 226–236. doi: 10.1177/0734282908330586

Lembke, E., and Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and first-grade students. *Learn. Disabil. Res. Pract.* 24, 12–20. doi: 10.1111/j.1540-5826.2008.01273.x

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav. Res. Methods* 48, 936–949. doi: 10.3758/s13428-015-0619-7

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

Meade, A. W., Johnson, E. C., and Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *J. Appl. Psychol.* 93, 568–592. doi: 10.1037/0021-9010.93.3.568

Milburn, T. F., Lonigan, C. J., DeFlorio, L., and Klein, A. (2019). Dimensionality of preschoolers' informal mathematical abilities. *Early Childhood Res. Quart.* 47, 487–495. doi: 10.1016/j.ecresq.2018.07.006

Muthén, L. K., and Muthén, B. O. (2017). *Mplus User's Guide*, 8th Edn. Los Angeles, CA: Muthén & Muthén.

National Council of Teachers of Mathematics (1989). *Curriculum and Evaluation Standards for School Mathematics*. Reston, VR: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*, Vol. 1. Reston, VR: National Council of Teachers of Mathematics.

Nelson, G., and McMaster, K. L. (2018). The effects of early numeracy interventions for students in preschool and early elementary: a meta-analysis. *J. Educ. Psychol.* 111, 1001-1022.

Nuutila, K., Tuominen, H., Tapola, A., Vainikainen, M.-P., and Niemivirta, M. (2018). Consistency, longitudinal stability, and predictions of elementary school students' task interest, success expectancy, and performance in mathematics. *Learn. Instr.* 56, 73–83. doi: 10.1016/j.learninstruc.2018.04.003

Parsons, S., and Bynner, J. (2005). *Does numeracy matter more?*. London: NRDC.

Polignano, J. C., and Hojnoski, R. L. (2012). Preliminary evidence of the technical adequacy of additional curriculum-based measures for preschool mathematics. *Assess. Effect. Interven.* 37, 70–83. doi: 10.1177/1534508411430323

Praet, M., Titeca, D., Ceulemans, A., and Desoete, A. (2013). Language in the prediction of arithmetics in kindergarten and grade 1. *Learn. Individ. Diff.* 27, 90–96. doi: 10.1016/j.lindif.2013.07.003

Purpura, D. J., Hume, L. E., Sims, D. M., and Lonigan, C. J. (2011). Early literacy and early numeracy: the value of including early literacy skills in the prediction of numeracy development. *J. Exp. Child Psychol.* 110, 647–658. doi: 10.1016/j.jecp.2011.07.004

Purpura, D. J., and Lonigan, C. J. (2013). Informal numeracy skills: the structure and relations among numbering, relations, and arithmetic operations in preschool. *Am. Educ. Res. J.* 50, 178–209. doi: 10.3102/0002831212465332

Purpura, D. J., and Lonigan, C. J. (2015). Early numeracy assessment: the development of the preschool early numeracy scales. *Early Educ. Dev.* 26, 286–313. doi: 10.1080/10409289.2015.991084

Purpura, D. J., Schmitt, S. A., and Ganley, C. M. (2017). Foundations of mathematics and literacy: the role of executive functioning components. *J. Exp. Child Psychol.* 153, 15–34. doi: 10.1016/j.jecp.2016.08.010

Purpura, D. J., and Simms, V. (2018). Approximate number system development in preschool: what factors predict change? *Cogn. Dev.* 45, 31–39. doi: 10.1016/j.cogdev.2017.11.001

Ribner, A., Moeller, K., Willoughby, M., and Blair, C. (2018). Cognitive abilities and mathematical competencies at school entry. *Mind Brain Educ.* 12, 175–185. doi: 10.1111/mbe.12160

Rittle-Johnson, B., Fyfe, E. R., Loehr, A. M., and Miller, M. R. (2015). Beyond numeracy in preschool: adding patterns to the equation. *Early Childhood Res. Quart.* 31, 101–112. doi: 10.1016/j.ecresq.2015.01.005

Rock, D. A., Pollack, J. M., and Hausken, E. G. (2002). *Early Childhood Longitudinal Study-Kindergarten Class of 1998-99 (ECLS-K): Psychometric Report for Kindergarten Through First Grade (NCES 2002-05)*, Washington, DC: National Center for Education Statistics, U.S. Department of Education.

Schmitt, S. A., Geldhof, G. J., Purpura, D. J., Duncan, R., and McClelland, M. M. (2017). Examining the relations between executive function, math, and literacy during the transition to kindergarten: a multi-analytic approach. *J. Educ. Psychol.* 109, 1120–1140. doi: 10.1037/edu0000193

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372

Toll, S. W. M., Van Viersen, S., Kroesbergen, E. H., and Van Luit, J. E. H. (2015). The development of (non-)symbolic comparison skills throughout kindergarten and their relations with basic mathematical skills. *Learn. Individ. Diff.* 38, 10–17. doi: 10.1016/j.lindif.2014.12.006

van Marle, K., Chu, F. W., Li, Y., and Geary, D. C. (2014). Acuity of the approximate number system and preschoolers' quantitative development. *Dev. Sci.* 17, 492–505.

Vernon-Feagans, L., Cox, M., Willoughby, M., Burchinal, M., Garrett-Peters, P., Mills-Koonce, R., et al. (2013). The family life project: an epidemiological and developmental study of young children living in poor rural communities. *Monogr.Soc. Res. Child Dev.* 78:150.

Conflict of Interest: 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.