Arithmetic Learning in Children: an fMRI Training Study

Merel Declercq[1], Elien Bellon[1], Muhammet Ikbal Sahan[2], Wim Fias[2], and Bert De Smedt[1]

[1]Department of Parenting and Special Education, KU Leuven

Leopold Vanderkelenstraat 32
B-3000 Leuven

[2]Department of Experimental Psychology, UGent

Henri Dunantlaan 2
B-9000 Gent

Abstract

Arithmetic learning is characterized by a change from procedural strategies to fact retrieval. fMRI training studies in adults have revealed that this change coincides with decreased activation in the prefrontal cortex (PFC) and that within the parietal lobe, a shift occurs from the intraparietal sulcus (IPS) to the angular gyrus (AG) during this change. It remains to be determined whether similar changes can be observed in children, particularly because children often recruit the hippocampus (HC) during fact retrieval, an observation that has not consistently been found in adults. In order to experimentally manipulate arithmetic strategy change, 26 typically developing 9- to-10-year-olds completed a six day at-home training of complex multiplication items (e.g. 16 x 4). Before and after training, children were presented with three multiplication conditions during fMRI: (1) complex to-be-trained/trained items, (2) complex untrained items and (3) single-digit items. Behavioral data indicated that training was successful. Similar to adults, children showed greater activity in the IPS and PFC for the untrained condition post-training, indicating that the fronto-parietal network during procedural arithmetic problem solving is already in place in children of this age. We did not observe the expected training-related changes in the HC. In contrast to what has been observed in adults, greater activity in the AG was not observed for the trained items. These results show that the brain processes that accompany the learning of arithmetic facts are different in children as compared to adults.

Keywords: fMRI, arithmetic, fact retrieval, learning, children

Arithmetic Learning in Children: an fMRI Training Study

A key aspect of children's cognitive development involves a change in the strategies they use to solve problems (e.g. Lemaire, 2017; Siegler, 2016). This has been frequently observed in the context of learning arithmetic (Bailey et al., 2012; Barrouillet et al., 2008; Geary et al., 1992; Lemaire & Siegler, 1995; Siegler & Shrager, 1984). Starting with simple counting strategies to solve small arithmetic items (e.g. 1, 2 and 3, 4, 5 to solve 2 + 3), children's arithmetic strategy use develops into more advanced counting strategies, such as counting on from the larger number (e.g. 3, 4, 5 to solve 2 + 3), potentially aided with finger representations (Geary et al., 1992). Later on, children learn to use even more advanced procedural strategies, where arithmetic problems are decomposed to be solved (e.g. 7 + 8 = 7 + 3 + 5 = 15, or 3 x 5 = 5 + 5 + 5 = 15). Through education, these small arithmetic items are repeated regularly, allowing for problem-answer associations to be formed. When these associations are stored in long-term memory, they are labeled as arithmetic facts (Siegler, 1996; Siegler & Shrager, 1984). The direct retrieval of these arithmetic facts is less time-consuming and more accurate than the abovementioned procedural strategies (Bailey et al., 2012). Although these arithmetic strategies and their development are researched well in children at the behavioral level, less is known about the neural changes that accompany this development (see Peters & De Smedt, 2018 for a review and Arsalidou et al., 2018 for a meta-analysis). Against this background, the current fMRI training study aimed to address three gaps in the literature. First, we tried to replicate the results of fMRI training studies in adults (e.g., Delazer et al., 2005; Grabner et al., 2009; Ischebeck et al., 2006, see Zamarian & Delazer, 2015 for a review) in a sample of children. Secondly, and in contrast to most of the existing body of adult data (however see Bloechle et al. (2016)), we added a pre-training fMRI assessment to fully map the neural changes that accompany the development of arithmetic strategies as a consequence of training. Finally, we aimed to further improve our knowledge of learning arithmetic in children by adding a condition of fully consolidated arithmetic facts to our fMRI design. This allowed us to investigate the differences in brain networks between already consolidated facts and the newly learned facts through training. These three aims are explained in more detail in the remainder of this introduction.

The majority of brain imaging studies on arithmetic have so far focused on adult participants (see Arsalidou & Taylor, 2011 for a meta-analysis). fMRI studies in adults have consistently found that performing a calculation activates a large fronto-parietal network, including the intraparietal sulcus (IPS) in the posterior parietal cortex (see Menon, 2015 for a review). The function of the IPS within arithmetic has been linked to the representation and manipulation of numerical magnitudes (Dehaene et al., 2003). Activation in the prefrontal cortex (PFC) often increases simultaneously with activity in the parietal cortex for arithmetic (see Menon, 2015 for a review). In an event-related fMRI study where adult participants had to process both correct and incorrect simple arithmetic equations, results showed that the activation of the PFC overlapped with brain areas known to be involved in working memory and interference processing (Menon et al., 2002). In a meta-analysis, Arsalidou & Taylor (2011) concluded that prefrontal activation during arithmetic acts as a general resource to cognitive functions such as working memory, and that these activations are based on the difficulty of the task. The fronto-parietal network is mainly active during large arithmetic problems (e.g. 42 + 74) and subtractions, which are most likely to be solved by using the procedural strategies, in which the numbers in the problem are being decomposed (e.g. 42 + 74 = 40 +70 + 2 + 4 = 116). Prefrontal activation is linked to these procedural strategies, and it is associated with working memory. Another set of activated regions for arithmetic can be seen in the angular gyrus (AG) and the supramarginal gyrus (SMG; Grabner et al., 2009; Tschentscher & Hauk, 2014). In contrast to the more difficult procedural problems that activate the fronto-parietal network, the AG and the SMG are activated for smaller arithmetic problems and for multiplication items. These items are more typically solved by retrieving the solutions from long-term memory (i.e. arithmetic fact retrieval).

The abovementioned shift from procedural strategies to arithmetic fact retrieval has been investigated at the neural level in adults using a training paradigm (e.g., Delazer et al., 2005; Grabner et al., 2009; Ischebeck et al., 2006, see Zamarian & Delazer, 2015 for a review). The shift from procedures to fact retrieval is examined by manipulating strategy use through an arithmetic training. When the items of the arithmetic training are difficult items that need to be solved via procedures at first, the training mimics the shift from using procedural strategies to retrieval from long-term memory by repeatedly exposing the participants to the problem-answer association. This training design combined with neuroimaging before and after the training allows one to manipulate brain-behavior relationships and to assess the impact of learning on brain activity patterns (Rosenberg-Lee, 2018). Training studies on arithmetic have been used extensively in adults (see Zamarian & Delazer, 2015 for a review). These training studies all have used complex arithmetic items (e.g. multi-digit multiplication items or artificial operations). Before and after the training, participants solve the items of the training in the scanner, as well as equally difficult untrained items. By comparing brain activity between the trained and untrained items, these studies are capable of showing which regions change as a function of training, and thus as a function of learning and transitioning from procedural strategies to arithmetic fact retrieval.

In one of the first fMRI training studies by Delazer et al. (2003), 13 adults trained 18 complex (double-digit x single-digit) multiplication problems daily over the course of a week. After the training, all participants underwent fMRI during which both the trained items and matched untrained items matched in difficulty were presented. When comparing brain activity during untrained items with trained items, greater activity in a large-scale network, including the IPS, the inferior parietal lobule and the inferior frontal gyrus was found. When comparing trained items with untrained items, greater brain activity was mainly observed in the left AG as well as the inferior temporal gyrus and the cingulate cortex, which is assumed to be activation associated with fact retrieval. These findings were replicated in a series of subsequent studies using similar training protocols with different manipulations (i.e. rote learning versus meaningful learning; Delazer et al., 2005; *N* = 16), or using addition rather than multiplication (Ischebeck et al., 2006; *N* = 17). A similar pattern of findings was also observed when using non-arithmetic tasks, such as a visuo-spatial learning task (participants had to learn how many planes a 3D object had; Grabner et al., 2009; *N* = 28). To the best of our knowledge, no such short-term training studies have been carried out in children. Thus, the first aim of the current study was to focus on how children learn arithmetic by replicating these arithmetic training studies in children. Such a study is important against the background of the observation that the brain activity during arithmetic in children shows substantial differences with that of adults (e.g., Peters & De Smedt, 2018). One example is that in children increases in brain activity during arithmetic fact retrieval are observed in the hippocampus (HC), whereas this is not the case in adults (Qin et al., 2014).

One critical limitation of existing adult fMRI training studies is that none of them measured brain activation before the training, as Rosenberg-Lee (2018) specifically advocated for. Not including an assessment before the arithmetic training is problematic because the mere comparison between trained and untrained items does not allow to evaluate the changes in brain activation associated with the actual learning of arithmetic facts, as is the case when comparing changes in brain activation before and after training. This issue was also addressed by Bloechle et al. (2016) in adults by adding an fMRI measurement before training in their arithmetic training study. Bloechle et al. (2016) reported greater activation for the untrained items compared with the trained items post-training in the IPS, and increased activation for the trained items compared with the untrained items post-training in the AG. These results are similar to the results of the training studies discussed above. However, by adding a pre-training scan, these authors were able to compare brain activity before and after training. For the trained versus to-be-trained contrast, Bloechle et al. (2016) could not replicate the increased AG activation, which has typically been found for the post-training analogue contrast. Instead, they showed that after training activation increased in a variety of regions, including the left HC, and bilateral SMG, insula and putamen. Bloechle et al. (2016) suggest that these results represent a central role of long-term memory in the medial temporal lobe for fact retrieval, rather than the AG. They hypothesized that the AG might serve as an interface that adjusts to attentional demands, acting as a circuit breaker. In this interpretation, the greater activity of the AG in arithmetic problem solving results from the AG ‘deciding’ whether a multiplication problem can be solved using arithmetic fact retrieval. If that is not the case, the response is to activate the fronto-parietal nework in order to increase cognitive resources. As Bloechle et al. (2016) showed with their pre- and post-training functional assessment, there are two ways of approaching the same concept (post-training comparison of trained and untrained problems versus the pre-post training fMRI comparisons of trained versus to-be-trained problems). Even though these are theoretically similar approaches, they do not result in the same findings. Additionally, a pre-training assessment provides a clear theoretical advantage in mapping the acquisition of arithmetic facts through training, as it allows for a full view of how the manipulation, in this case the training, changed brain activity (Rosenberg-Lee, 2018). Thus, our second aim of this study was to include a pre-training measurement to compare the same items before and after the training.

Although there are similarities in brain activation during arithmetic between adults and children, as for example in the recruitment of fronto-parietal areas, there are also differences, especially for the role of the medial temporal lobe, the activation of which is particularly observed during arithmetic fact retrieval (Qin et al., 2014). More specifically, De Smedt et al. (2011) found that in 10-to-12-year-old children, a fronto-parietal network including the IPS was recruited for large problems and for subtractions, similar as for adults. These types of more difficult items are more likely to be solved using procedural strategies. On the other hand, diverging results between children and adults are observed for problems that are solved through fact retrieval. In adults, greater activity in the AG have been observed in small arithmetic addition problems (i.e. single-digit items < 10). De Smedt et al. (2011), however, showed that in 10-12-year-old children, greater activity in the left HC was observed for items more likely to be solved using fact retrieval. In a more recent study with children of a similar age, Polspoel et al. (2017) classified each item as either solved by procedure or retrieval using a trial-by-trial strategy assessment outside the scanner. This design allowed for a very precise assessment of the strategy that the participants used. This is a necessary step, as previous studies described above infer strategy use solely from item characteristics such as ‘small versus large’ or ‘addition versus subtraction’ when focusing on fact retrieval versus procedural strategies. In adults, it has been shown that the strategy used to solve an arithmetic items determines neural activation rather than the arithmetic operation that is being presented (Fresnoza et al., 2020; Tschentscher & Hauk, 2014). Polspoel et al. (2017) used an event-related design to classify each item individually as being solved via fact retrieval or via procedural strategy, based on the strategy assessment outside the scanner. Thus, brain activity could be analyzed for procedure or retrieval items for each participant individually. Retrieval strategies were accompanied by increased activity in the bilateral AG and the SMG, while the procedural strategies showed increased activation in the IPS and in prefrontal regions. These results are in line with what has been observed from adults’ cross-sectional and training studies, yet they are somewhat different from what has been observed in children of the same age (De Smedt, et al., 2011; Qin et al., 2014).

The abovementioned studies are important to unravel the neural activation during arithmetic in children. However, one crucial shortcoming of these studies is that they are unable to investigate development in a more detailed manner, as they only investigate one age group Focusing more on development, Rivera et al. (2005) studied children of a large age range (ages 8 to 19). Analyzing the brain activity during single-digit addition and subtraction, they showed that while older children showed increased activation in the left parietal cortex and left lateral occipital-temporal cortex, younger children mainly showed greater activation in the PFC, the HC and the dorsal basal ganglia. Prado et al. (2014) studied arithmetic fluency in children from second grade to seventh grade (ages 8 to 14). They observed grade-related increases of activity in the mid temporal gyrus for multiplication. Prado et al. (2014) concluded that the increased reliance on mid temporal regions during development is caused by the increased reliance on arithmetic fact retrieval.

In another approach to study development, Cho et al. (2011) classified 7-to-9 year-old children as either retrievers or counters depending on their retrieval frequency (i.e. when more than 60% of the items were categorized as either retrieval or counting in a verbal trial-by-trial a single-digit addition task). Cho et al. (2011) assumed that the retrievers were further along their arithmetic development as compared to the counters and consequently that the comparison of these two groups of children had the potential to reveal something about children’s development towards retrieval. Univariate analyses revealed greater responses in the ventro-lateral PFC for retrievers, while multivariate analysis revealed distinct activity patterns in the bilateral HC, bilateral posterior parietal cortices and left ventro-lateral PFC between retrievers and counters. In a follow-up analysis, Cho et al. (2012) showed that higher retrieval fluency (i.e. the percentage of trials correctly retrieved in a single-digit addition verbal strategy assessment) was correlated with greater activation in the right HC, parahippocampal gyrus and other regions.

Finally, one longitudinal study has investigated the development of arithmetic in the brain at multiple time points. Qin et al. (2014) studied the role of the HC in the transition from procedures to retrieval. This was done by testing 8-year old children again at age 9, as well as comparing these data with cross-sectional adolescent and adult data. Strategy assessment of the addition problems used in this study showed more frequent retrieval use at age 9 compared to the first time point at age 8. This increased frequency of retrieval use continued into adolescence and adulthood. Qin et al. (2014) also demonstrated that the bilateral HC showed increased activity during addition between the two childhood time points, while the activation in the left superior parietal lobule the bilateral dorso-lateral PFC decreased. Additionally, children at age 9 had the largest hippocampal activity, also compared to adolescents and adults. These observations suggested a time-limited role for the HC in the development of arithmetic. The studies mentioned above focus on development by either studying different age groups, with changes in retrieval frequency as a marker for development (e.g., Prado et al., 2011; Rivera et al., 2005), or by using a longitudinal design (e.g., Qin et al., 2014). This developmental approach is not the same as focusing on the learning process and the accompanying changes in brain activity that occur over a shorter amount of time. This would require a design with a manipulation of the learning process, as is done in a training study, over a short amount of time (see Rosenberg-Lee, 2018, for a similar critique), as we will do in the current study.

Taken together, the abovementioned studies in children point towards a decrease of activation in the PFC and parietal cortex and an increase in the hippocampal activity during the development of arithmetic. An important difference between adults and children is the involvement of the HC rather than the AG in arithmetic retrieval in children. Only a few studies in children have reported increased activation in the AG (Polspoel et al., 2017) as has been observed in adults (Delazer et al., 2003, 2005; Grabner et al., 2009; Ischebeck et al., 2009; Zamarian & Delazer, 2015). On the other hand, increases in hippocampal activation have been repeatedly found in children, (Cho et al., 2011, 2012; Peters & De Smedt, 2018; Polspoel et al., 2017; Prado et al., 2014; Qin et al., 2014; Rivera et al., 2005), while in adults it was only found in a training study when considering brain activity before and after the training (Bloechle et al., 2016). These divergences might occur because the role of HC is time-dependent. During learning, the HC might aid the consolidation of arithmetic facts into long-term memory. After these facts have been consolidated, their retrieval is being carried out by the AG, and the role of the HC diminishes (see Peters & De Smedt, 2018, for a discussion). Based on the abovementioned equivocal results, it is crucial to further disentangle the role of the HC and AG during arithmetic fact retrieval. This was precisely the third aim of the current study. One way to test this differential contribution of HC and AG, is to compare the brain activity of newly learned facts, to the brain activity of already well-consolidated facts. The hypothesis is that for the newly learned facts, the HC will show greater activation, while for the already consolidated facts, the AG is assumed to play a larger role. We operationalized this third aim by adding a third condition of consolidated arithmetic facts to the traditional trained and untrained conditions of our arithmetic training study. This third condition included very well-known items (e.g. 3 x 5) that were fully consolidated in the children of this study. By comparing these fully consolidated arithmetic facts with the trained items of the training, we can investigate time-dependent effects that occur during the learning of arithmetic facts. This allows us to disentangle the role of the HC and AG during the acquisition of arithmetic facts.

Current Study

In the current study we used an experimental arithmetic training paradigm to study arithmetic learning in children, simulating their change in strategy use from procedure use to fact retrieval. This approach has been applied in adults with success (see Zamarian et al., 2015 for a review) but has never been applied to children using fMRI. Specifically, we designed a training study that lasted for six days, during which double-digit x single-digit multiplication items (e.g. 14 x 3) were trained in a sample of 10 year-old children. This narrow age range was chosen in order to minimize potential maturational and educational effects. Many of the existing body of data included children of a broader age range, yet by doing so, these studies introduce variables, such as age and educational history, that are hard to fully control for. These broad age ranges are valuable for studies focusing on development, but to be able to capture the learning of arithmetic facts as precisely as possible research focused on a specific age should also be considered. In light of that, this study included children who were all very close in age. The reason for specifically focusing on 10-year-olds was twofold. Firstly, all participants needed to be fully proficient in the single-digit multiplication items. This was needed because we included a condition of single-digit arithmetic items that was expected to be fully automatized in order to realize the third aim of our study. In the Belgian educational system, the multiplication tables up to 10 are taught in the second year of primary school, and are expected to be fully known by heart from then on. Secondly, participants needed to understand the concept of multiplication well enough such that they had the skills to solve more complex multiplication items during the training, namely the double-digit multiplication items of the trained and untrained conditions.

We specifically focused on children because findings from adult studies, which can inform us about the organization of mature cognitive processes, cannot be readily generalized to children (Ansari, 2008). One example of this is the potential role of the HC versus AG in arithmetic fact retrieval, which is predicted to be different in children and adults (e.g., Peters & De Smedt, 2018). Furthermore, even if behavioral performance in children is comparable to that of adults (e.g. both adults and children are able to solve problems via fact retrieval), this does not necessarily imply equivalence in neurocognitive mechanisms (Ansari, 2008). Against this background, it is valuable to also investigate how arithmetic strategies change in children and what neural processes accompany these changes.

By experimentally manipulating the change from procedural use to fact retrieval through training, we aimed to make three important contributions to the existing knowledge on the neural correlates of arithmetic in children.

Our first aim was to investigate if the results of the fMRI training studies in adults could be replicated in children. This was done by comparing trained and untrained items post-training. We hypothesized that the untrained items would show greater activation compared to the trained items in a frontoparietal network, with important roles for the IPS and the PFC, reflecting the use of procedural strategies for untrained arithmetic items. Against the background of the existing developmental data (Cho et al., 2011, 2012; Qin et al., 2014), we hypothesized that the trained items would show greater activation in the HC, as these items are trained to form arithmetic facts.

Our second aim was to include a pre-training fMRI measurement, following the recommendations by Bloechle et al. (2016) and Rosenberg-Lee (2018). This allowed us to better explore the changes caused by the training, rather than relying on the more traditional, post-training contrasts only. We hypothesized that the multiplication items before training compared to after training would show increased activity in a network including the IPS and PFC, similar to adults. We also hypothesized that the trained items compared with the same items before training, would show increased activity in the HC, and decreased activity in the fronto-parietal network (Bloechle et al., 2016).

Our third aim, which comprised the largest theoretical novelty, was to test the potential time-dependent roles of the HC and AG in learning arithmetic. This was done by comparing the trained items, i.e. the newly learned multi-digit items of the training with items that had been already automatized and consolidated by children of this age, i.e. single-digit multiplication items. Research in adults suggested that activity in the AG is more reflective of fact retrieval of well-consolidated, automated arithmetic facts. On the other hand, developmental data indicate that greater activity in the HC are expected for newly learned items. We therefore expected that the trained items would show increases in activity in the HC while the fully automatized items (single-digit items) would show increased activation in the AG.

Methods

Participants

Participants were 45 typically developing Flemish 4th and 5th graders, who were recruited via their elementary schools. None of them had a history of learning difficulties or any neurological disorders. Each participant as well as a parent of each child signed a written informed consent form. The study was approved by the Medical Ethical Committee of the KU Leuven (S59167). For each participated session, the child received a financial compensation.

Procedure

The study design included several steps, namely (1) a behavioral assessment of arithmetic, reading and intellectual ability, which also included a familiarization with the mock MRI scanner, (2) a pre-training fMRI assessment, (3) an at-home training for six consecutive days, and finally (4) a post-training fMRI assessment the day after the last training day. Seven participants were too scared to go in the scanner after the mock scanning session, while four children in the pre-training scan, five in the post-training scan, and three in both sessions were excluded due to excessive head movement (see motion criteria below). As we only included participants with a full dataset, this resulted in a final sample of 26 children (M_{age} = 10.4 years, *SD* = 0.34,13 girls).

Pre-training Behavioral Assessment

During the pre-training session, participants were tested on arithmetic (Tempo Test Arithmetic, TTA; De Vos 1992), reading (One-minute Test, OMT and Klepel; Van den Bos et al., 1994) and intellectual abilities (i.e., verbal and performance intelligence measured with similarities and block design subtests from the WISC-III-NL; Wechsler, 2005) to characterize their academic performance and ability level. Standardized scores were calculated for all these tasks, which are shown in Table 1. Participants had average to above average scores on arithmetic, reading and intellectual abilities. We also tested whether the excluded participants scored significantly different on these behavioral measures (independent *t*-tests), which was not the case for any of the four measures (TTA, *t*(43) = -0.92, *p* = .36, OMT *t*(43) = 0.95, *p* = .35, Klepel *t*(43) = 0.03, *p* = .98, verbal intelligence *t*(43) = 0.28, *p* = .78 and performance intelligence *t*(43) = - 0.13, *p* = .90).

Table 1

Scores for Arithmetic, Reading, Verbal, and Performance Intelligence of the Pre-training Behavioral Assessment

Competence	Test	Mean standardized score (*SD*)	Range
Arithmetic	Tempo Test Arithmetic (TTA; De Vos, 1992)	6.5 (3.0)	1 – 10
Reading	One-minute test (OMT) &	10.4 (3.0)	6 – 16
	Klepel (Van den Bos et al., 1994)	11.9 (2.5)	5 – 19
Verbal intelligence	Similarities, WISC-III-NL NL (Wechsler, 2005)	14.0 (2.4)	9 – 17
Performance intelligence	Block design, WISC-III-NL NL (Wechsler, 2005)	12.0 (3.2)	6 – 17

Note. Scores for arithmetic are standardized as *M* = 5 and maximum of 10. Scores for reading, verbal intelligence and performance intelligence are standardized as *M* = 10 and a maximum of 20.

To test whether the double-digit multiplication items used in the training (see below, Table 2) were not known as arithmetic fact by the participants prior to the training, they all performed a verbal trial-by-trial strategy assessment task. During this assessment, children solved the 40 items one by one. After each item, the children were asked to explain how they solved the problem, encouraging them to use any strategy that was available to them. Each answer was categorized as either fact retrieval (e.g. which they indicate by “I know this problem from class” or “it was already in my head”), procedure (e.g. repeated addition, decomposition, solving a neighboring multiplication and correcting for that) or other (if the child was unable to explain his solving strategy properly or when the explanation was not clear). Of the 40 items in total, 20 were the double-digit multiplication items, 10 of set 1 and 10 of set 2 (see below, Table 2). We additionally included 20 already known arithmetic facts to contrast these new items with. These single-digit multiplication items were related to the two training sets (e.g. 6 x 4 was included because 16 x 4 is part of one of the training sets).

Training

Training Sets. For this study, two sets of 10 double-digit multiplication items were designed (see Table 2) to be used during the training. Both sets consisted of 10 two-digit (range 12-19) times one-digit (range 3-8) problems with solutions between 42 and 98, with the first operand always the two-digit number. The two sets were matched on size of the solution ($M_{\text{set 1}}$ = 75.9, *SD* = 18.7, $M_{\text{set 2}}$ = 73.1, *SD* = 16.3, *t*(18) = -0.357, *p* = .73), on size of the two-digit number ($M_{\text{set 1}}$ = 15.5, *SD* = 2.3, $M_{\text{set 2}}$ = 14.9, *SD* = 2.5, *t*(18) = -0.560, *p* = .58), and on size of the one-digit operand ($M_{\text{set 1}}$ = 5.0, *SD* = 1.5, $M_{\text{set 2}}$ = 5.1, *SD* = 1.7, *t*(18) = 0.142, *p* = .89). Problems that would result in tie-problems if solved through decomposition were excluded (e.g. 16 x 6). Half of the participants received set 1 as the training set, and set 2 as the untrained set in the fMRI, while the other half of the participants received set 2 as the training set and set 1 as the untrained set in the fMRI to balance out any set-related differences.

Table 2
Two Sets of 10 Multiplication Items for Training

Set 1		Set 2	
12 x 7 = 84	13 x 6 = 78	12 x 6 = 72	12 x 8 = 96
14 x 3 = 42	14 x 7 = 98	13 x 5 = 65	13 x 7 = 91
15 x 6 = 90	16 x 4 = 64	14 x 6 = 84	15 x 4 = 60
17 x 3 = 51	17 x 5 = 85	16 x 3 = 48	17 x 4 = 68
18 x 4 = 72	19 x 5 = 95	18 x 5 = 90	19 x 3 = 57

Note. Half of the participants trained set 1 and received set 2 as the untrained set, while the other half trained set 2 and received set 1 as the untrained set during fMRI acquisition.

Training Protocol. Similar to training studies in adults (Grabner et al., 2009), participants completed a training of one of the developed training sets on six consecutive days at home, for approximately 20 minutes a day. The parents of the participants were asked not to provide any help, to make sure no help in any other form was used, and were encouraged to let the training happen in a quiet environment. Each participant was randomly assigned either set 1 or set 2. Each training day, the 10 assigned problems were presented in a block in random order, with each training day consisting of 10 blocks. One training day thus consisted of 100 items. At the start of each training day, children were instructed to respond as quickly and as accurately as possible. As the training lasted for 6 days, and as each item was presented 10 times on each day, each item was trained 60 times. All problems were presented on a black screen in a white font. The answers were given using the computer keyboard, and response time was defined from the start of the item until the confirmation of the answer with the ENTER key. Depending on the accuracy of the response, the participant either got positive or negative feedback. Before being able to move on to the next item, the problem and its correct solution were presented on the screen for two seconds, in order to stimulate the correct memory association between a problem and its solution. To encourage the transition to fact retrieval, the response time was limited to 12 seconds in the first 2 training days, to 10 seconds in days 3 and 4, and to 8 seconds in days 5 and 6. If an answer was not correct, or not provided within the given timeframe, feedback was given and the correct solution was presented for two seconds before the next trial could begin. To familiarize participants with the number keypad, each training started with the copying of 10 randomly selected double-digit numbers. A pilot study (*N* = 12) in children of the same age (grade 4) showed that the training was successful in decreasing response times and increasing accuracy after six days of training. Note that the same pilot study showed that for children of one grade lower, i.e. grade 3 (8-9 years), the double-digit multiplication items were too difficult to carry out the training and to undergo the quick-paced fMRI protocol adequately.

Imaging

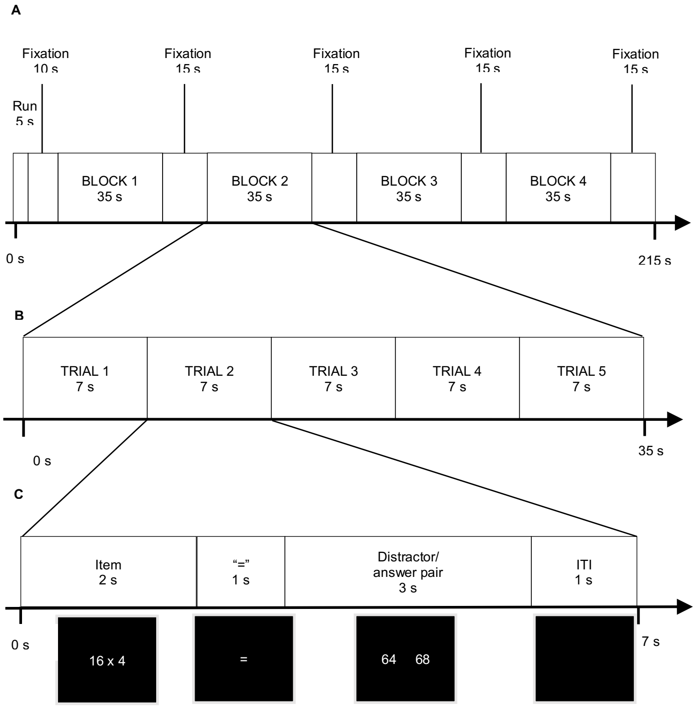
fMRI: Experimental Design. Both the pre-training and post-training fMRI test session used the exact same protocol and design. During scanning, participants solved multiplication items belonging to one of three experimental conditions. The three experimental conditions consisted of (1) the 10 trained double-digit times single-digit multiplication items (TR, the items from the assigned training set), (2) the 10 untrained double-digit times single-digit multiplication items (UN, the items from the other set) and (3) 10 related single-digit times single-digit multiplication items (single-digit related, SR, related to the assigned training set). Each condition from the pre-training is denoted with ‘1’ (i.e. TR1) and each condition from the post-training session will be given a ‘2’ (i.e. TR2).

In addition to the three experimental conditions, we also included a control condition. This condition showed a two-digit number and a one-digit number, separated by a ‘#’ (e.g. 81 # 3, visually similar to the trained and untrained condition). Participants had to remember the two numbers, and were presented with one new number and one number from the trial (e.g. 91 and 3). They had to press left if the two-digit number was the same as from the trial, or right if the one-digit number was similar. Because initial analyses revealed that this task relied too heavily on working memory, the data of this condition was not further analyzed.

Stimuli were presented in a block design, which consisted of six runs (see Figure 1). A run consisted of three experimental blocks (TR-UN-SR) and the control block, which was always acquired at the end of the run. The order of the experimental blocks per run was counterbalanced to have a unique order of experimental conditions for each run. Each run started with the run number for five seconds and a 10-second fixation before starting with a trial. Within a block, five trials of the same condition were grouped in a block, lasting for 35 seconds. Blocks were separated by 15 seconds inter-block intervals, showing a fixation cross. Stimuli were presented with E-prime 3.0 software (Psychology Software Tools, Pittsburgh, PA) for two seconds in white font on a black background, followed by an equal sign for one second. After that, two potential answers were shown for three seconds in a delayed verification paradigm, in which participants had to indicate the side of the correct answer by left or right button press. The next trial started after a one second inter-trial interval. The position of the correct answer was left-right balanced, as well as the type of distractor that was provided. The distractor was either operand-related, with half of the items related to the first operand and the other half of the second operand, both plus or minus 1, or solution-related, with the distractor being either 10 units bigger or smaller than the correct solution. The type of distractor was balanced between all six possibilities (e.g. first operand + or – 1; second operand + or – 1; solutions + or – 10), and all six possibilities were left/right balanced.

Figure 1

Schematic Overview of In-scan Paradigm



Note. Schematic view of one run (A), one block (B), and one trial (C). One run was repeated six times for each participant in each session.

MRI Data Acquisition and Analysis. Functional and structural images were acquired by a Philips Ingenia 3.0T CX MRI scanner with a SENSE 32- channel head-coil, located at the

Department of Radiology of the University Hospital in Leuven, Belgium. Structural imaging was performed between run three and four. Scanning time per session was around 35 minutes. To minimize head motion, participants' heads were stabilized. For the fMRI data, 52 slices were recorded in an ascending order, using a T2*-sequence (2.19 x 2.19 x 2.2 mm voxel size, 2.2 mm slice thickness, 0.3 mm interslice distance, 96 x 95 acquisition matrix, 90° flip angle) and covered the whole brain (field of view: 210 x 210 x 130 mm). Each run consisted of 73 images (TR = 3000 ms, TE = 29.8 ms). Anatomical images were acquired with a T1-weighted sequence (0.98 x 0.98 x 1.2 mm voxel size, 256 x 256 acquisition matrix, 8° flip angle, TE 4.6 ms, 250 x 250 x 218 mm field of view).

All preprocessing steps were conducted with the Statistical Parametric Mapping (SPM12, Wellcome Department of Cognitive Neurology, London) software using Matlab for pre-training and post-training sessions separately. All functional images were spatially realigned to the mean functional image, and were then temporally realigned to the middle slices. The realigned images were then coregistered to the anatomical image, and then normalized to the standard Montreal Neurological Institute (MNI) space. Finally, the normalized functional images were smoothed with a spatial filter of 8 mm full-width at half maximum Gaussian smoothing kernel. This larger smoothing kernel was chosen as this study had a young population, where (too much) motion is of a larger concern compared to adults. Although several measures were taken in order to minimize motion (i.e. stabilization of the head), a larger smoothing kernel was chosen than usually recommended to account for this issue. We excluded runs with excessive motion (i.e. if the movement of from one image to the next was greater than one voxel size). If a participant had less than four out of six usable functional runs per session, he or she was excluded from the analyses. Although 45 children participated, analysis was performed on 26 participants (M_{age} = 10.4 years, *SD* = 0.34,13 girls) because of fear to go in the MRI (*n* = 7), or excessive motion in the pre-training scan (*n* = 4), in the post-training scan (*n* = 5), or in both scans (*n* = 3).

A general linear model was calculated, which combined pre- and post-training scans for each participant in one model. For each session, the three experimental conditions and the control condition were modeled as boxcar function for the duration of the block, as well as all fixations. The motion realignment parameters were included as regressors of no interest. As runs were concatenated, run regressors of no interest for each run that was included were also added in the model as well as a regressor for the pre-training and one for the post-training fMRI session. First-level contrasts were calculated for each participant, and then combined on group level using a one-sample *t*-test. Significant clusters were labeled using the anatomical automatic labeling tool (AAL, Rolls et al., 2015) in SPM 12. Second-level results were assessed using a family-wise error (FWE) correction for multiple comparisons **at the voxel level** of *p* <

.05. Only clusters of 20 or more voxels are reported. The *t*-contrasts were calculated without any masking.

Results

Behavioral Results

Pre-training Assessment

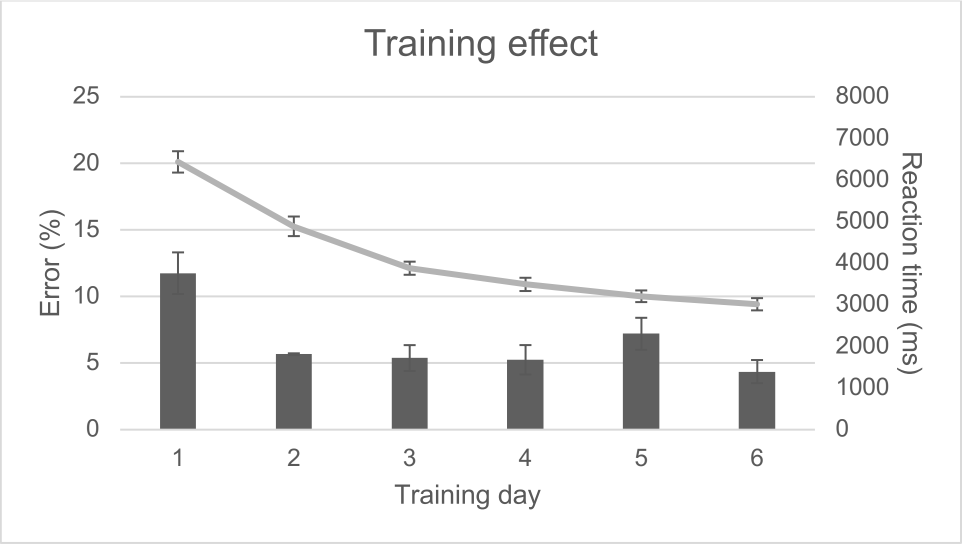
The verbal trial-by-trial pre-training assessment revealed that participants were able to solve the complex double-digit multiplication items of the training sets (M_{acc} = 90 %, *SD* = 8.3, M_{rt} = 7222 ms, *SD* = 2416). Analyses of the verbal protocols confirmed that for the double-digit multiplication items, the procedural strategy was the most dominant (*M* = 98.4 %, *SD* = 5.5). This confirms that the items that were to be trained (TR1) were not yet stored as arithmetic facts, and consequently this allows us to study the effects of training.

For the single-digit multiplication items, accuracy was higher and response time faster than for the double-digit multiplication items (M_{acc} = 97 %, *SD* = 3.8, M_{rt} = 2967 ms, *SD* = 1543). As predicted, most single-digit multiplication items were solved using arithmetic fact retrieval (*M* = 93.2 %, *SD* = 11.5).

Training

All 26 participants completed the training as described above, but for two participants, a technical error occurred and results could only be recovered for two training days each. Training effects were thus analyzed for 24 participants (Figure 2). For both accuracy and response time, a repeated measures ANOVA was performed with training day as a within-subject factor. For accuracy, a statistical significant effect of training was found ($F(5, 115) = 15.26$, $p < .001$, $\eta^2 = 0.399$). Bonferroni-corrected post-hoc *t*-tests showed a significant difference between day one and two ($t(23) = -4.5$, $p = .002$), but not between the other days. A significant effect of training for response time was also found ($F(1.8, 42.3) = 214.7$, $p < .001$, $\eta^2 = 0.903$). Bonferroni-corrected post-hoc *t*-tests showed a statistical significant difference between each consecutive training day: Between day one and two ($t(23) = 14.3$, $p < .001$), two and three ($t(23) = 8.6$, $p < .001$), three and four ($t(23) = 6.3$, $p < .001$), four and five ($t(23) = 4.9$, $p < .001$), and between five and six ($t(23) = 4.1$, $p = .007$). As there are clear training effects, the training was considered to have worked.

Figure 2
Behavioral Effects of the Arithmetic Training



Note. Accuracy expressed in error rate (%, left y-axis, bars) and response times (ms, right y-axis, line graph). Error bars indicate ±1 SE of the mean, *N* = 24

In-scan Behavioral Results

As the in-scanner task (choice reaction) deviated from the training task outside the scanner (production), we additionally analyzed the participants' in scanner performance. First, we investigated whether the participants were able to give an answer within the given time frame of the scanning protocol. Second, we checked whether accuracy was above chance level. Lastly, we verified if children had improved on the training items at the post-training scan. Descriptive statistics for these variables can be found in Table 3.

Table 3
Average Scores for the In-scan Behavioral Results

Variable		**Trained**		**Untrained**		**Single-digit**	
		Pre	**Post**	**Pre**	**Post**	**Pre**	**Post**
Too late	%	8	3	10	10	2	3
	SD	8	5	11	10	3	4
On time, correct	%	74	90	70	74	95	94
	SD	17	10	18	13	7	9
On time, incorrect	%	18	5	20	16	3	3
	SD	16	6	16	10	7	5
Response time	ms	1498	1191	1577	1612	1073	1104
	SD	222	236	308	316	142	158

Note. Response time was only calculated for correctly solved items. Response time was measured from the onset of the choice possibilities.

First, the results in Table 3 show that participants were able to give an answer within the given time frame for the majority of trials, both in the pre-training and post-training sessions, and for all three conditions. As one could expect, single-digit multiplication items were most often answered within the provided time (98% and 97% pre- and post-training respectively), while untrained multiplication items showed some lower percentages (90% for both the pre- and post-training scan). The trained multiplication items showed similar percentage to the untrained items before training (92%), but after the training, the amount of problems solved on time for the trained items resembled more the single-digit multiplication items (97%). Of the answers given in time, the single-digit multiplication items were answered correctly most of the time (95% and 94% pre- and post-training respectively). The untrained items' accuracy of the answers that were given on time increased slightly between the two scans (70% and 74% pre- and post-training respectively), while for the trained items, the increase was more pronounced (74% and 90% pre- and post-training respectively).

Secondly, one-sample t-tests were conducted to determine if the accuracy of the three conditions both pre- and post-training statistically differed from the chance level of 50%. Accuracy was significantly different from chance level accuracy for all the conditions, both pre- and post-training (TR1 ($t(25) = 7.099$, $p < .001$), TR2 ($t(25) = 23.080$, $p < .001$), UN1 ($t(25) = 5.883$, $p < .001$), UN2 ($t(25) = 9.290$, $p < .001$), SR1 ($t(25) = 31.382$, $p < .001$) and SR2 ($t(25) = 25.248$, $p < .001$)).

Finally, we evaluated whether the participants improved both on accuracy and for response time for the in-scan functional tasks by performing a 2 (time; pre-training scan versus post-training scan) x 3 (condition; TR versus UN versus SR) repeated measures ANOVA. For accuracy, there was a significant main effect of time ($F(1, 25) = 12.75$, $p = .001$, $\eta^2 = .060$), a significant main effect of condition ($F(2, 50) = 70.56$, $p < .001$, $\eta^2 = .476$), and a significant interaction effect between time and condition ($F(2, 50) = 23.93$, $p < .001$, $\eta^2 = .086$). Bonferroni-corrected post-hoc tests showed that when comparing the same condition pre- and post-training, there were significant differences for the to-be trained versus the trained condition ($t = -7.056$, $p < .001$), but not for the untrained ($t = -1.510$, $p = 1.000$) and single-digit conditions ($t = 0.530$, $p = 1.000$). Pre-training, there was no significant difference between the trained and untrained condition ($t = 1.548$, $p = 1.000$), but post-training, they significantly differed ($t = 7.324$, $p < .001$), with the trained condition being more accurate. Pre-training, the to-be-trained condition differed significantly with the single-digit condition ($t = -9.088$, $p < .001$), but post-training, they did not differ significantly anymore ($t = -1.198$, $p = 1.000$). There were significant differences between the untrained and the single-digit both pre-training ($t = -10.636$, $p < .001$) and post-training ($t = -8.522$, $p < .001$). For response time, there was a significant main effect of time ($F(1,25) = 13.23$, $p = .001$, $\eta^2 = .023$), a significant main effect of condition ($F(2, 50) = 82.31$, $p < .001$, $\eta^2 = .609$) and a significant interaction effect between time and condition ($F(2, 50) = 49.58$, $p < .001$, $\eta^2 = .092$). Bonferroni-corrected post-hoc tests showed that when comparing the same condition pre- and post-training, there were significant differences for the to-be trained versus the trained condition (*t* = 9.680, *p* < .001), but not for the untrained (*t* = -1.090, *p* = 1.000) and single-digit conditions (*t* = -0.990, *p* = 1.000). Pre-training, there was no significant difference between the trained and untrained condition (*t* = -1.790, *p* = 1.000), but post-training, they significantly differed (*t* = -9.552, *p* < .001). Pre-training, the to-be-trained condition differed significantly with the single-digit condition (*t* = 9.645, *p* < .001), but post-training, they did not differ significantly anymore (*t* = 1.954, *p* = .817). There were significant differences both pre-training (*t* = 11,435 *p* < .001) and post-training (*t* = 11.507, *p* < .001) between the untrained and the single-digit items. Taken together, the changes of the in-scanner performance on the trained items from pre- to post-training, which were not observed for the untrained and single-digit items, confirm that our training protocol was effective.

Imaging Results

This imaging results section is organized according to the three aims: (1) The replication of the adult studies by comparing the trained and untrained multiplication items post-training, (2) The investigation of the training effect in children by comparing the double-digit multiplication items of the trained condition before and after training and (3) The examination of the role of the HC versus the role of the AG in children by comparing, both before and after training, the single-digit multiplication items with the double-digit multiplication items. In addition, we report four supplementary contrasts, grouped in two supplementary analyses. The first two contrasts focus on the comparison of the to-be-trained and untrained conditions before the start of the training. These two conditions are qualitatively similar conditions of unknown double-digit multiplication items before training. The next two contrasts are the untrained items before and after training. These were calculated to verify whether the training of items with very similar characteristics yielded a transfer effect that impacts the solving of the untrained items.

For some of the contrasts, we had very specific anatomical hypotheses, i.e. hypotheses on the IPS, SMG, AG and the HC. In the case that the results at the corrected threshold did not support these hypotheses, we ran an exploratory analysis by looking at more fine-grained anatomical ROIs. These four ROIs were anatomically defined through the Anatomy Toolbox (Eickhoff et al., 2007), using the cytoarchitectonic maps and the subdivisions they provide. For the HC, we included CA1, CA2 and CA3 left and right, for the AG we included PGa and PGp left and right, for the SMG we included PF, PFm, PFcm, PFt and PFop left and right, and for the IPS we included hIP1, hIP2 and hIP3 left and right. To control for multiple comparisons in these exploratory analyses, we opted for a small-volume correction using these ROIs.

Comparing Untrained (UN2) and Trained Items (TR2) of the Post-training Session (aim 1)

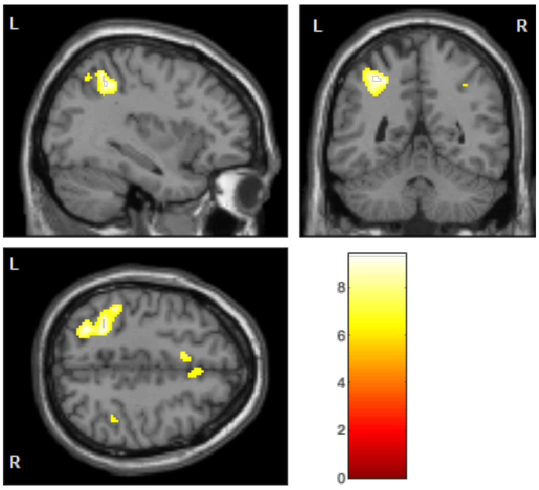
Here, we discuss the results coupled with our first aim, i.e. the replication of the adult studies by comparing the trained and untrained multiplication items post-training.

UN2-TR2. The contrast of untrained and trained double-digit x single-digit multiplication items at the post-training session showed significant greater activation in clusters throughout the whole brain, including greater activation in the inferior parietal lobule, particularly in the left hemisphere. There was also greater activation for the untrained items compared to the trained items in the left cerebellum, SMA, precentral gyrus and inferior temporal lobule. There was greater activation for the untrained items compared to the trained items in the right superior frontal gyrus and inferior frontal operculum (see Figure 3, and Table 4 for coordinates).

TR2-UN2. The trained double-digit multiplication items yielded no significantly greater activation compared to the untrained double-digit multiplication items post-training. For this contrast, we hypothesized greater activation in the HC for the trained items. We therefore repeated the analysis with the small-volume correction for the hippocampal subdivisions (CA1, CA2 and CA3), yet this exploratory analysis did not yield any significant differences.

Figure 3

Visualization of the Comparison of Untrained (UN2) and Trained (TR2) items Post-training



Note. UN2-TR2 (x = -34, y = -50, z = 48). TR2-UN2 showed no suprathreshold clusters. Statistical threshold was $p < .05$, corrected for multiple comparison using a family-wise error (FWE) correction at the voxel level. Only clusters of 20 or more voxels are reported.

Table 4
Region, Peak Coordinates, Cluster Size (k), and t-values for the Comparison of Untrained (UN2) and Trained (TR2) Items at the Post-training Session

Cluster	Peak coordinates			k	*t*
	x	y	z		
UN2-TR2					
L inferior parietal lobule	-34	-50	48	782	9.39
L inferior parietal lobule	-42	-40	46		
L superior parietal lobule	-28	-40	46		
L cerebellum	-6	-76	-30	121	9.34
L SMA	-6	12	52	371	8.99
R mid cingulate cortex	8	18	42		
R medial superior frontal cortex	4	26	42		
L precentral gyrus	-42	2	32	264	8.48
R superior frontal gyrus	28	8	56	72	7.55
R inferior frontal operculum	52	16	34	35	6.79
L inferior temporal lobule	-48	-56	-16	21	6.71
R inferior parietal lobule	40	-42	46	36	6.69

Note. Statistical threshold was $p < .05$, corrected for multiple comparison using a family-wise error (FWE) correction at the voxel level. Only clusters of 20 or more voxels are reported. Cluster labeling was done via the AAL tool (Rolls et al., 2015). Minor maxima of clusters larger than 300 are also reported.

Comparing To-be-trained (TR1) and Trained Items (TR2; aim 2)

The following results discuss the second aim of our study and compare the double-digit multiplication items before and after training.

TR1-TR2. The to-be-trained multiplication items of the pre-training fMRI session yielded no significantly greater activation compared to the trained multiplication items of the post-training fMRI session. For this contrast, we hypothesized greater activation in the IPS for the pre-training items. We therefore repeated the analysis with the small-volume correction for the IPS' subdivisions (hIP1, hIP2 and hIP3), yet this exploratory analysis did not yield any significant differences.

TR2-TR1. The trained multiplication items of the post-training fMRI session yielded no increased activation compared to the to-be-trained multiplication items of the pre-training session. For this contrast, we hypothesized increased activation in the HC for the post-training items. We therefore repeated the analysis with the small-volume correction for the hippocampal subdivisions (CA1, CA2 and CA3), yet this exploratory analysis did not yield any significant differences.

Comparing Double-digit Items (TR and UN) with Single-Digit items (SR) Before and After Training (Aim 3)

In the following two sections, we discuss our third aim by comparing the single-digit multiplication items with the double-digit multiplication items. Because this comprised eight comparisons in total, we have divided the results further into two sections. In the first section, we report the comparisons of both the trained and untrained multiplication items with the single-digit multiplication items. In the second section, we report the reverse comparisons.

Double-digit Items (TR and UN) versus Single-digit items (SR).

TR1-SR1. Pre-training, there was greater activity in the to-be-trained multiplication items compared to the single-digit multiplication items in the bilateral SMA, in the right SMG and superior occipital lobule, and in the left inferior parietal lobule, cerebellum, mid frontal lobule, inferior occipital lobule and inferior temporal lobule (see Figure 4A, and Table 5 for coordinates).

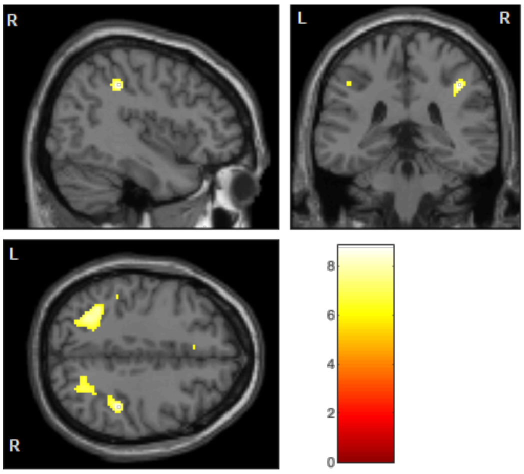
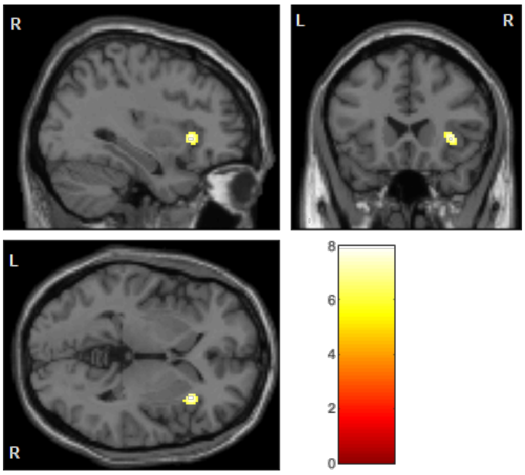
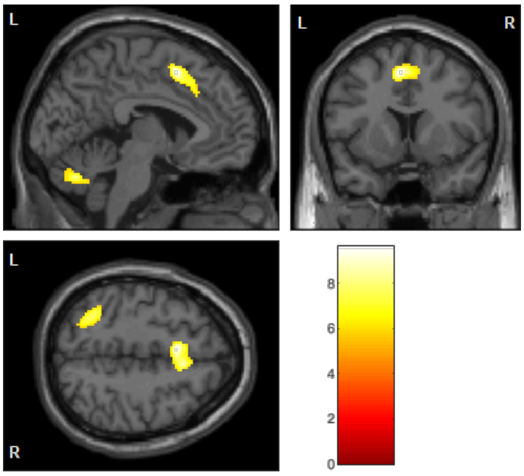
UN1-SR1. Pre-training, there was significant greater activation in the right insula in the untrained items compared to the single-digit items (see Figure 4C, and Table 5 for coordinates). For this contrast, we hypothesized greater activation in the IPS for the untrained items. We therefore repeated the analysis with the small-volume correction for the IPS' subdivisions (hIP1, hIP2, hIP3), yet this exploratory analysis did not yield any significant differences.

TR2-SR2. Post-training, there was no significant difference in activation between the trained items compared to the single-digit items. For this contrast, we hypothesized greater activation in the HC for the trained items. We therefore repeated the analysis with the small-volume correction for the hippocampal subdivisions (CA1, CA2 and CA3), yet this exploratory analysis did not yield any significant differences.

UN2-SR2. Post-training, there was significantly greater activation in the bilateral cerebellum, in the right insula, and in the left SMA, inferior frontal lobule, precentral gyrus, inferior parietal lobule and fusiform gyrus in the untrained items compared to the single-digit items (see Figure 4D, and Table 5 for coordinates).

Figure 4

Visualization of the Comparison of Double-digit (TR and UN) and Single-digit Items (SR) Both Pre- and Post-training

	PRE-TRAINING	**POST-TRAINING**
TR – SR	**A. TR1-SR1** 	**B. TR2-SR2** No suprathreshold clusters
UN - SR	**C. UN1-SR1** 	**D. UN2-SR2** 

Note. A. TR1-SR1 (x = 42, y = -36, z = 42), B. TR2-SR2, C. UN1-SR1 (x = 34, y = 22, z = 00) and D. UN2-SR2 (x = -4, y = 10, z = 52). Statistical threshold was $p < .05$, corrected for multiple comparison using a family-wise error (FWE) correction at the voxel level. Only clusters of 20 or more voxels are reported.

Table 5

Region, Peak Coordinates, Cluster Size (k), and t-values for the Comparison of Double-digit (TR and UN) and Single-digit Items (SR) both at Pre- and Post-training Session

	PRE-TRAINING						**POST-TRAINING**					
	Region	**Peak coordinates**			**k**	***t***	**Region**	**Peak coordinates**			**k**	***t***
		x	**y**	**z**				**x**	**y**	**z**		
TR – SR	R supramarginal gyrus	42	-36	42	127	8.84	No suprathreshold clusters					
	L inferior parietal lobule	-38	-50	52	727	8.39						
	L superior parietal lobule	-30	-62	48								
	L inferior parietal lobule	-34	-52	42								
	L cerebellum	-30	-62	-28	170	8.31						
	Left mid frontal lobule	-26	2	54	33	7.74						
	L inferior occipital lobule	-22	-94	-4	48	7.68						
	L SMA	-8	8	60	34	7.46						
	R SMA	10	14	48	100	7.38						
	L inferior temporal lobule	-50	-54	-12	28	7.03						
	R superior occipital lobule	28	-64	42	335	6.93						
	R precuneus	20	-66	44								
	R angular gyrus	34	-56	46								

	Region	Peak coordinates			k	t	Region	Peak coordinates			k	t
		x	y	z				x	y	z		
UN – SR	R insula	34	22	0	46	7.97	L SMA	-4	10	52	648	9.63
							R mid CC	10	20	40		
							R SMA	6	16	52		
							L inferior frontal lobule	-48	26	30	189	9.03
							L precentral gyrus	-40	4	34	155	8.31
							L inferior parietal lobule	-34	-48	42	654	8.26
							L inferior parietal lobule	-30	-56	52		
							L superior parietal lobule	-30	-62	46		
							L cerebellum	-4	-70	-30	206	8.16
							L fusiform gyrus	-38	-78	-14	155	7.42
							R insula	38	18	-4	28	7.00
							R cerebellum	28	-64	-28	84	6.68

Note. Statistical threshold was *p* < .05, corrected for multiple comparison using a family-wise error (FWE) correction at the voxel level. Only clusters of 20 or more voxels are reported. L = left, R = right, SMA = supplementary motor area. Cluster labeling was done via the AAL tool (Rolls et al., 2015). Minor maxima of clusters larger than 300 are also reported.

Single-digit Items (SR) versus Double-digit Items (TR and UN).

SR1-TR1. Pre-training, the single-digit items showed greater activity than the to-be-trained items in the left mid temporal lobule and medial frontal orbital lobule (see Figure 5A, and Table 6 for coordinates). For this contrast, we hypothesized greater activation in the HC and the AG for the single-digit multiplication items. We therefore repeated the analysis with the small-volume correction for the hippocampal subdivisions (CA1, CA2 and CA3) and AG subdivisions (PGa and PGp), yet this exploratory analysis did not yield any significant differences.

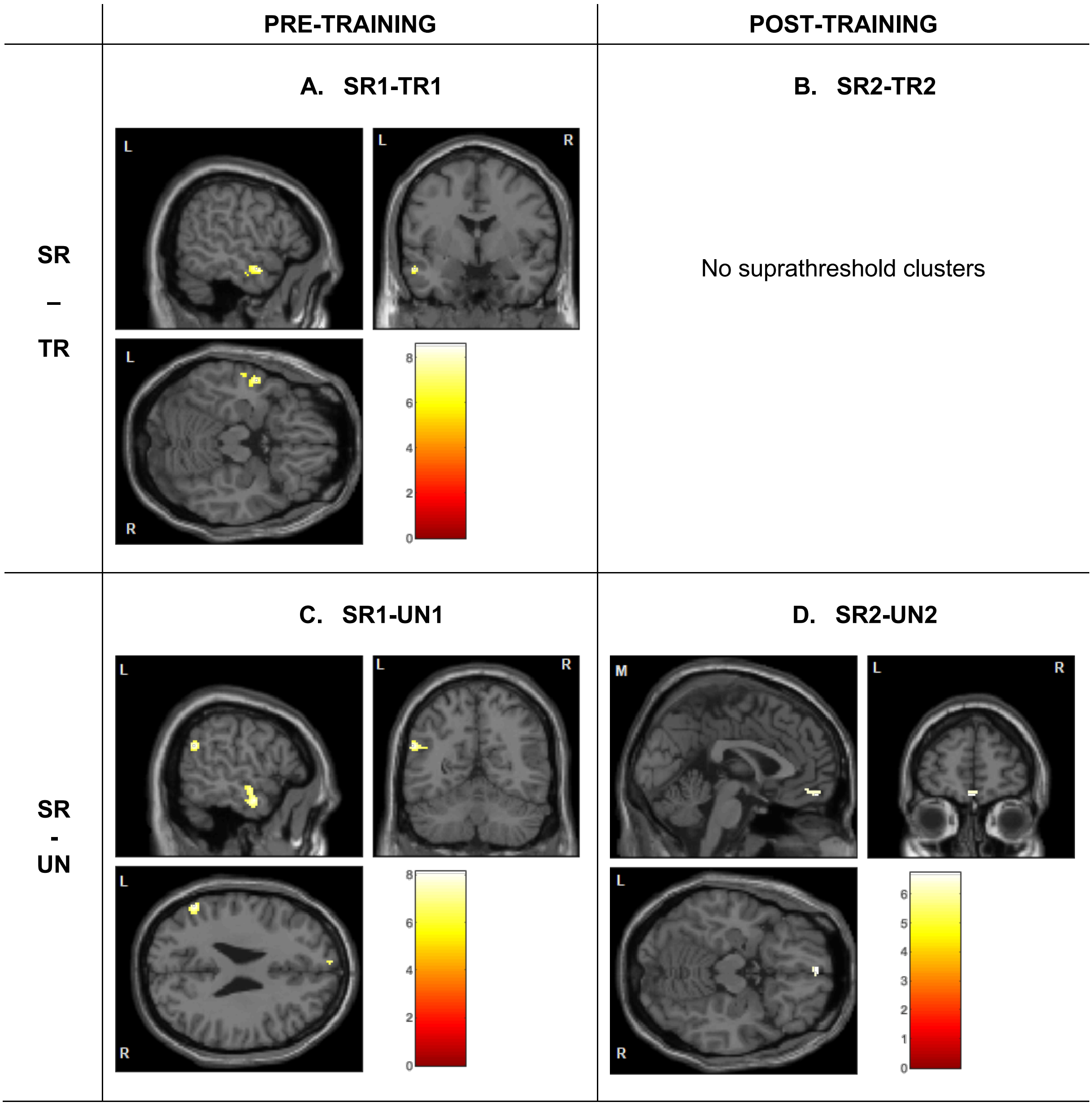
SR1-UN1. Pre-training, the single-digit items showed greater activity than the untrained items in the left AG, mid temporal lobule, frontal superior medial lobule and paracentral lobule, and in the right medial frontal orbital lobule (see Figure 5C, and Table 6 for coordinates). For this contrast, we hypothesized greater activation in the HC for the single-digit multiplication items. We therefore repeated the analysis with the small-volume correction for the hippocampal subdivisions (CA1, CA2 and CA3), yet this exploratory analysis did not yield any significant differences. It is important to emphasize that the differences observed in the AG were differences in activation and not less de-activation.

SR2-TR2. Post-training, there was no significant greater activation in the single-digit items compared to the trained items. For this contrast, we hypothesized greater activation in the HC and the AG for the single-digit multiplication items. We therefore repeated the analysis with the small-volume correction for the hippocampal subdivisions (CA1, CA2 and CA3) and AG subdivisions (PGa and PGp), yet this exploratory analysis did not yield any significant differences.

SR2-UN2. Post-training, the single-digit items showed greater activation than the untrained items in the left rectus (see Figure 5D, and Table 6 for coordinates). For this contrast, we hypothesized greater activation in the HC and the AG for the single-digit multiplication items. We therefore repeated the analysis with the small-volume correction for the hippocampal subdivisions (CA1, CA2 and CA3) and AG’s subdivisions (PGa and PGp), yet this exploratory analysis did not yield any significant differences.

Figure 5

Visualization of the Comparison of Single-digit Items (SR) and Double-digit (TR and UN) both Pre- and Post-training



Note. A. SR1-TR1 (x = -54, y = -2, z = -18), B. SR2-TR2, C. SR1-UN1 (x = -56, y = -58, z = 26) and D. SR2-UN2 (x = 0, y = 56, z = -16). Statistical threshold was $p < .05$, corrected for multiple comparison using a family-wise error (FWE) correction at the voxel level. Only clusters of 20 or more voxels are reported.

Table 6

Region, Peak Coordinates, Cluster Size (k), and t-values for the Comparison of Single-digit Items (SR) and Double-digit (TR and UN) Items Both at Pre- and Post-training Session

	PRE-TRAINING						**POST-TRAINING**					
SR – TR	**Region**	**Peak coordinates**			**k**	***t***	**Region**	**Peak coordinates**			**k**	***t***
		x	**y**	**z**				**x**	**y**	**z**		
	L mid temporal lobule	-54	-2	-18	138	8.59	No suprathreshold clusters					
	L medial frontal orbital lobule	0	56	-12	82	7.05						
SR – UN	**Region**	**Peak coordinates**			**k**	***t***	**Region**	**Peak coordinates**			**k**	***t***
		x	**y**	**z**				**x**	**y**	**z**		
	L angular gyrus	-56	-58	26	75	8.13	L rectus	0	56	-16	35	6.72
	R medial frontal orbital lobule	4	54	-8	225	7.93						
	L mid temporal lobule	-56	-4	-22	272	7.85						
	L frontal superior medial lobule	-8	50	8	25	7.76						
	L frontal superior medial lobule	-4	62	24	30	7.74						
	L paracentral lobule	-2	-28	50	21	6.80						

Note. Statistical threshold was *p* < .05, corrected for multiple comparison using a family-wise error (FWE) correction at the voxel level. Only clusters of 20 or more voxels are reported. L = left, R = right. Cluster labeling was done via the AAL tool (Rolls et al., 2015).

Supplementary analyses

Baseline Before Training. Here, we discuss the results of the comparison of the to-be-trained and untrained items before training. As at this point, the two conditions should not be qualitatively different for which reason we expect no significant differences between these two conditions.

UN1-TR1. Pre-training, there were no significant differences in activation between the untrained items and the to the to-be-trained items.

TR1-UN1. Pre-training, there were no significant differences in activation between the to-be-trained items and the untrained items.

Transfer Effect in the Untrained Multiplication Items. Here, we discuss the results of the comparison of the untrained items before and after training. Even though the untrained items are not part of the training between scans, it might be possible that training multiplication items with very similar characteristics, as was carried out in the training condition, results in a transfer effect.

UN1-UN2. There was no significant greater activation in the untrained items before the training compared to the same items after the training.

UN2-UN1. There was no increased activation in the untrained items after the training compared to the same items before the training.

Discussion

This short-term training study of arithmetic multiplication items was designed to manipulate arithmetic learning and strategy change from procedural use to fact retrieval and to investigate the corresponding changes in brain activity. fMRI short-term training designs have been applied in a wide range of studies in adults already (Delazer et al., 2003, 2005; Grabner et al., 2009; Ischebeck et al., 2006, 2007), but they have not been used to study arithmetic learning in children. Earlier brain imaging studies in children have used either cross-sectional (Cho et al., 2011, 2012; Prado et al., 2014; Rivera et al., 2005) or longitudinal designs (Qin et al., 2014), yet they did not investigate the effect of an experimental manipulation of arithmetic strategy use via training on brain activity. The current study is the first to apply such a training design in children. This allowed us to investigate the brain activity during trained and untrained items after training and to verify if findings derived from adult arithmetic training studies replicate in children. To address a critical limitation of the existing adult training studies (Bloechle et al., 2016; Rosenberg-Lee, 2018), the current study also included a pre-training scan via which we could directly investigate the changes in brain activity. Lastly, and against the background of developmental studies that have suggested a time-dependent role of the HC during arithmetic fact retrieval (Cho et al., 2011, 2012; De Smedt et al., 2011; Prado et al., 2011; Qin et al., 2014; Rivera et al., 2005), we contrasted the brain activity of the newly trained items with an additional third experimental condition of well-known arithmetic facts. This allowed us to investigate the difference between newly learned and already consolidated arithmetic facts in children.

In the present study, 9-to-10 year-old children trained double-digit times single-digit multiplication items (e.g. 16 x 4) for six days. This age group was chosen because of their proficiency in single-digit arithmetic and their adequate conceptual understanding of multiplication, via which they should be able to solve more complex double-digit multiplication items. A narrow age range was chosen in order to minimize educational and maturational effects. A pilot study showed that children from one grade lower (8-9 years old) had too many difficulties with double-digit multiplication items to carry out the training or to undergo the quick-paced fMRI protocol adequately. Before and after the training, brain activity was assessed for the items in the trained set, the items in the untrained set, and single-digit multiplication items. Our first aim was to investigate children's brain responses to learning new arithmetic items and their similarities and differences with what has been observed in adults. Against the background of the available literature, we hypothesized that the untrained items would show increased activation compared to the trained items in a fronto-parietal network, and thus that there would be little differences between children and adults for this contrast. We hypothesized that the trained items would show increased activation in the HC because of the hypothesized time-dependent role of the HC in children, rather than the AG as has been observed in adults. For the second aim, we examined the changes before training and after training for the to-be-trained/trained items to unravel the changes that arithmetic learning causes in children. We hypothesized that the multiplication items before training compared to after the training would show increased activity in a similar network to that in adults, including the IPS and PFC. We also hypothesized, different from what has been observed in adults, that the trained items compared with the same items before training, would show increased activity in the HC, and decreased activity in the fronto-parietal network, parallel to the hypothesis for the post-training contrast of the first aim. Finally, for our third aim, we investigated the time-dependent role of the HC during arithmetic learning by comparing two different conditions of arithmetic fact retrieval (i.e. fully consolidated single-digit items versus newly learned double-digit multiplication items). We hypothesized that increased activation in the AG would be found for mature fact retrieval of the well-consolidated arithmetic facts, while the newly trained multiplication items would show increased activation in the HC. We discuss our findings in relation to these three aims below and end our discussion by elaborating on the study limitations and future directions.

Trained and Untrained Items at the Post-training fMRI (aim 1)

To realize our first aim, we examined two contrasts: Untrained items at the post-training fMRI minus trained items at the post-training fMRI and its reverse contrast. The results concerning the untrained items compared to the trained items (UN2-TR2) generally replicated findings observed in adult training studies. Untrained items produced greater activation in the IPS, predominantly on the left, which has most often been interpreted to reflect quantity-based processing that is necessary for the procedural solving of arithmetic items (Dehaene et al., 2003). The involvement of the frontal cortex for the untrained items next to the IPS is also a consistent finding, as is evidenced by the meta-analysis of Arsalidou and Taylor (2011). This meta-analysis on fMRI data on arithmetic and number tasks suggests that the involvement of the inferior frontal cortices is to process information that requires cognitive operations in order to attain task solution. Finally, for the untrained items, we also found greater activation in the left cerebellum, the left SMA, and the left inferior temporal lobe. Although meta-analyses have shown that increases in activation in these regions during arithmetic are often found both in adults (Arsalidou and Taylor, 2011) and in children (Arsalidou et al., 2018), these results are only occasionally discussed in detail. In their meta-analysis, Arsalidou and colleagues propose that the role of the cerebellum in mathematics is most likely related to the coordination of visual motor sequencing, particularly when the participants are under time constraints (Arsalidou & Taylor, 2011). Motor functions are known to activate the cerebellum, which strengthens this interpretation. The potential role of the SMA in arithmetic might be related to the fact that the SMA has been associated with movement sequencing (Nieder & Dehaene, 2009) and finger counting during mental calculation in children (Berteletti & Booth, 2015). Previous literature has suggested that the SMA relates to finger perception when subjects engage in arithmetic problem solving (Yang et al., 2017), and the current results might be reflective of this.

It is important to note that these reverse inference interpretations of brain activity should be considered carefully. Although many of the fMRI studies on arithmetic have tried to explain the involvement of clusters such as the cerebellum, insula, or SMA in specific arithmetic processes, most of these regions have also been found in studies on other cognitive processes (Yarkoni et al., 2012; Wang et al., 2010). It is important to stress this point, as reverse inference, or the use of reasoning from activation to mental functions, has been criticized (Poldrack, 2011). We use this approach as a “reasoning to the best explanation”, not to state facts that are set in stone or that they are the only interpretation of activity in these regions (Poldrack, 2011). This admonition on reverse inference should be considered throughout the rest of the discussion.

The second contrast for our first aim examined the trained items compared to the untrained items post-training (TR2-UN2). Different from our expectations, we could not replicate the results of the adult studies, namely the greater activation in the AG for trained multiplication items. We also could not could confirm the hypothesis for greater HC activation in the trained items, as was predicted on the basis of existing data in children (Qin et al., 2014).

The additional exploratory ROI-analyses with small-volume correction did not show any significant effects and consequently did not confirm our hypotheses.

One possible explanation for the lack of activity in the AG in this study might come from suggestion that the AG acts as an attention allocator that plays a role in the strategy selection decision process (Bloechle et al., 2016). According to this hypothesis, the AG is active whenever a person needs to decide to either retrieve the correct solution from long-term memory, or to either calculate the solution using procedural strategies. As the current study used a block design, items that elicited the same strategy were grouped together within one block, which probably did not result in a lot of demands on attention allocation to select a particular strategy. This need for attention allocation to select the most appropriate strategy would mostly exist in an event-related design. In such a design there would be a larger need to select and switch between strategies for each item individually. Interestingly, Polspoel et al. (2017) used an event-related design to investigate arithmetic strategies during multiplication in children of a similar age range. These authors indeed observed increases in AG during fact retrieval in children, which strengthens the hypothesis of an attention allocator role for the AG. Future studies are needed to test this possibility by adapting an event-related design with sequences where a strategy decision needs to be made every trial, and sequences where strategy is stable of the course of several trials.

Additionally, and in contrast to our hypothesis based on literature in children (e.g. De Smedt et al., 2011; Prado et al., 2011), we also could not find greater hippocampal activation for the trained items. One possible explanation of why we could not confirm our hypothesis is that in general, the number of significant clusters in the trained versus untrained contrast is much less numerous than for the untrained versus trained contrast (Grabner et al., 2009; Ischebeck et al., 2009). This might be reflective of an overlap in solution strategy for untrained and trained items. As the pre-training strategy assessment showed, most of the double-digit multiplication items are solved using decomposition strategies, which involve retrieving intermediate solutions from memory (e.g. for solving 16 x 4, 10 x 4 and 6 x 4 are easily retrieved arithmetic facts as intermediate solutions). These intermediary retrieval steps within a procedural strategy most likely also activate regions that are correlated with fact retrieval in the brain, given that intermediate facts are retrieved. As a result, the use of arithmetic fact retrieval during the decomposition strategy makes it harder to fully reveal the training effect on trained items compared to untrained items, as both of the conditions include the retrieval of arithmetic facts.

Pre versus Post-training (aim 2)

Although thoroughly investigating the post-training results is important, earlier studies showed that is also critical to consider the pre- versus post-training effects (Bloechle et al., 2016), as also recommended by Rosenberg-Lee (2018). This allows one to directly to see training effects, as this consists of a comparison of the same items before and after training.

Neither of the two contrasts included in this aim (TR2-TR1 and TR1-TR2) showed any significant greater activation. Different from what has been found in adults, we could neither replicate decreases in a broader fronto-parietal network for TR1-TR2 nor increases in brain activity in the HC for TR2-TR1 (Bloechle et al., 2016).

We hypothesized an increase in the HC as a result of increased automatization after training for two different contrasts, i.e. trained versus untrained items post-training of the first aim (TR2-UN2), and trained items after training versus the same items before training of the second aim (TR2-TR1). For both contrasts, the results did not confirm our hypotheses, even when looking at the HC specifically as an ROI and applying a small volume correction. As discussed above, this might be because of the inherent use of fact retrieval in the most applied decomposition strategy also activates the retrieval network, masking differences in neural activity. Another possibility for the lack of greater HC activation might be that the duration of the training in children was not long enough to induce training-related changes in brain activity, particularly not in the HC. However, the current training protocol was longer and more intensive than the adult training studies (Grabner et al., 2009; Ischebeck et al., 2006, 2007) and the data at the behavioral level showed that training indeed was sufficient. We observed a clear behavioral decrease over all training days for response time, i.e. from seven seconds to three seconds, and accuracy quickly reached ceiling. Such a pattern of data can only be seen if the training was indeed successful in learning new items to a point where they are being automatized. Successful training effects were also seen when analyzing the in-scan accuracy scores and response times. For the trained items, accuracy significantly improved between the two fMRI sessions, while response times significantly decreased. This was neither the case for the untrained items nor for the single-digit multiplication items. As behaviorally the training was successful, it is plausible that the brain-related changes, for example in the HC, were not captured in the single post-training fMRI assessment. On the one hand, it might be needed to expand the duration of the training or to postpone the post-training fMRI assessment for a few days in order to see the expected changes in brain activity in children. On the other hand, it could be that the training-related changes in the HC are brief and very time-specific. For example, Van Opstal et al. (2008) observed that during the training of ordered sequences in adults activity in the HC already changed after one scanning session. It might therefore be interesting for future studies to include intermediate scanning sessions as well as scans with a longer interval post-training to fully map these neural changes.

Consolidated Items Compared To Trained And Untrained Items (aim 3)

Our third aim was to further understand role of the AG and the HC in children's arithmetic learning. This was done by adding a third condition of single-digit multiplication items, which were fully consolidated and are retrieved from long term memory, next to the trained and untrained condition. This resulted in eight contrasts which are discussed as follows: We first present the results of contrasts of the double-digit multiplication items minus the single-digit multiplication items, starting with the pre-training (i.e. TR1-SR1 and UN1-SR1) and proceeding with the post-training contrasts (i.e. TR2-SR2 and UN2-SR2). Then, we discuss the reverse contrasts in the same order (i.e. SR1-TR1, SR1-UN1, SR2-TR2 and SR2-UN2).

Before training, the to-be-trained multiplication items showed greater activation compared to the single-digit multiplication items (TR1-SR1) in the bilateral SMA, in the right SMG and superior occipital lobule, and in the left inferior parietal lobule, the mid frontal lobule, the inferior occipital lobule and the inferior temporal lobule. This greater activation in the fronto-parietal network supports complex arithmetic, as has been observed in both adults (see Menon, 2015 for a review) and in children (see Peters & De Smedt, 2018 for a review), as discussed above (see UN2-TR2).

The to-be-trained condition (TR1) and the untrained condition (UN1) were similar conditions, as they are both complex double-digit multiplication items. As such, it is surprising that the two contrasts with single-digit multiplication items (i.e. TR1-SR1 and UN1-SR1) were not similar to each other. The UN1-SR1 contrast only showed greater significant activation in the right insula. The additional exploratory ROI analysis with small-volume correction in the IPS did not yield significant differences, even though such differences were expected and observed in the TR1-SR1 contrast. In a supplementary analysis, we examined the baseline before training by comparing to-be-trained multiplication items (TR1) with the untrained condition (UN1). Both contrasts (i.e. UN1-TR1 and TR1-UN1) showed no significant differences between the conditions. However, the difference between the results of the pre-training contrasts comparing the double-digit items with the single-digit multiplication items (i.e. TR1-SR1 and UN1-SR1) did not entirely confirm that they were indeed fully equal conditions. Further exploration of this phenomenon in the data showed that this might have been the case because of the small sample size, combined with unbalanced drop-out creating an inequality between the number of participants that had set 1 as the trained set and the amount of participants that had set 2 as the trained set. Initially, the 45 participants were divided such that the distributions were 22 to 23, but after drop out, this was now 9 to 17. We performed a series of sanity checks to examine the effect of this inequality. First, we evaluated whether the drop-out group significantly differed from the group that was included, based on their training results (accuracy and response times) and on their the behavioral measures (reading, arithmetic, verbal and performance intelligence), but this was not the case. Secondly, we tested whether within the included group, their training results and their scores on the behavioral measures differed between the participants who were assigned set 1 as trained set, or set 2, but again they did not differ significantly. It is important to emphasize that these data are all behavioral data (production task) and that the response modality in the scanner was slightly different (verification task), due to the restrictions of the scanner environment. We therefore cannot exclude the possibility that this difference impacted on the imbalance between the sets in the scanner.

For the trained items post-training, there were no clusters with significant greater activation compared to the single-digit items (TR2-SR2). After the training, we expected the trained items to be solved using a similar strategy as the single-digit items, namely arithmetic fact retrieval. These results show that the differences between trained and single-digit items decrease as a result of training. However, we expected the trained and single-digit items to still be different, as the trained items are new items that are not yet fully consolidated compared to the single-digit items. For this TR2-SR2 contrast specifically, we expected the HC to show increased activation, but even using a small-volume correction, we could not confirm this hypothesis. Despite this evidence at the behavioral level that the training had induced its expected effects, it might be possible that, as already discussed above, the timing of the post-training assessment came too early to observe the expected differences in the HC in children.

Looking at the contrasts of the post-training session of the untrained minus the single-digit items (UN2-SR2), we observed the hypothesized greater activation in the untrained items compared to the single-digit multiplication items in the fronto-parietal network, with greater activation in the bilateral cerebellum, in the right insula, and the left SMA, inferior frontal lobule, precentral gyrus, inferior parietal lobule and fusiform gyrus. As already discussed above, this is similar to the activation that supports complex arithmetic problems, as has been observed in both adults (see Menon, 2015 for a review) and in children (see Peters & De Smedt, 2018 for a review), and as also seen in the contrasts between the untrained items minus the trained items post-training (UN2-TR2) and the trained items minus the single-digit items pre-training (TR1-SR1).

We now focus on the reverse contrasts of aim 3 (i.e. single-digit items versus double digit items). When comparing the single-digit items with the to-be-trained items pre-training (SR1-TR1) greater activation was found in the left middle temporal and medial frontal lobule. This is in line with previous studies as grade-related increases of activity in the left middle temporal lobe in primary school children have been observed in children from 2nd to 7th grade (Prado et al., 2014). These grade-related changes have been interpreted as being reflecting arithmetic fact retrieval and the involvement of verbal processes related to such fact retrieval, and the current differences between the single-digit items and double-digit items might be reflective of that.

For the similar contrast of the single-digit items minus the untrained items pre-training (SR1-UN1), we observed greater activation for the left AG, together with greater activation in the middle temporal lobule. This is expected, as the single-digit multiplication items are well consolidated in the participants (Delazer et al., 2005; Qin et al., 2014; Rosenberg-Lee et al., 2011, 2018). These clusters have been associated with fact retrieval, which is also the expected strategy for the single-digit multiplication items. Participants' use of fact retrieval is further supported by the strategy assessment data before the training: 93.2 % of the single-digit items were reported to be solved using arithmetic fact retrieval. Increases in left lateralized activation in the AG is a common finding in adults studies (Delazer et al., 2003, 2005; Grabner et al., 2009; Ischebeck et al., 2006, 2007), and in some studies on arithmetic in children (Polspoel et al., 2017), and these activation increases have been interpreted as reflective of arithmetic fact retrieval.

Post-training, there was no significant greater activation in the single-digit multiplication items compared to the trained items (SR2-TR2). Together with reverse contrast (i.e. TR2-SR2), this suggests that in the two conditions children use similar strategies, and therefore call upon similar neuronal resources to solve the multiplications.

Because the untrained items were not practiced during the training, we expected to observe that the single-digit versus untrained items contrast of the pre-training assessment (SR1-UN1) remained the same post-training (SR2-UN2). This is not what we found: Greater activity was only significant in the left rectus gyrus for this contrast. The untrained items were repeated three times during the acquisition of the fMRI data and it might be that through this very small repetition of items that was part of the testing protocol, children may have started to learn these untrained items. The untrained items were also visually and arithmetically close to the items of the trained set, which was done to make the trained and untrained as similar as possible. It is plausible that the learning of the trained set might have helped children to learn the items of the untrained set, or primed them to also try to learn these multiplication items. Such transfer effects might have been larger in children than in adults, a possibility that needs further investigation. In all, this might have resulted in the differences between the single-digit and untrained condition post-training to diminish, making it more difficult to distinguish between the single-digit items and the untrained items after training. If this was the case, one should also have seen differences pre- versus post-training in the untrained condition. However, no such direct transfer effects were observed in the supplementary analyses of these contrasts (UN1-UN2 and UN2-UN1), for which reason this hypothesis of differential transfer should be interpreted with great caution.

Taken together, the current pattern of behavioral and brain imaging data showed that the training was sufficient to change brain activation, and yet, we did not find the hypothesized potential time-dependent changes in the HC. In contrast to our predictions, we did not see increases in hippocampal activity in the trained items as a function of training, as has been observed in adults (Bloechle et al., 2016). We also did not observe increases in hippocampal activity when new learned multiplications (i.e. trained items) and already consolidated multiplications (i.e. single-digit items) were contrasted. However, the contrast of consolidated (single-digit) multiplications to the untrained items pre-training revealed increases in AG activity, a finding that is consistent with what has been observed in adults (e.g., Delazer et al., 2005; Grabner et al., 2009)) and in some developmental studies (Polspoel et al., 2017). Yet on the other hand, we did not see this for the trained items versus single-digit items pre-training and we did not see any differences in AG activity post training, for which reasons the findings related to the AG should be interpreted with great caution.

This observation, together with the inability to replicate the activation of the AG in adults in the trained items post-training, seems to suggest that brain activity as a consequence of learning arithmetic facts in the AG is different for children than for adults. At the moment, several hypotheses exist in how the involvement of the AG supports arithmetic fact retrieval. Originally, it was hypothesized that the left AG activity during arithmetic fact retrieval reflects language-mediated processes, i.e. the retrieval of arithmetic facts from verbal memory (Dehaene et al., 2003; Grabner et al., 2009). There is indeed evidence to suggest that more efficient arithmetic fact retrieval is related to verbal processes, such as phonological awareness (De Smedt et al., 2010), which could explain the role of these language-mediated processes. In more recent work, Bloechle et al. (2016) suggested that the role of AG reflects an attentional process that is related to the selection of strategies rather than to the actual retrieval of arithmetic facts as already discussed in more detail above. Another hypothesis attributes the AG activity to the automatic mapping between arithmetic problems and their solutions into long-term memory (Ansari, 2008). This ties together with the research where the AG is seen as part of the semantic network. For example, neural activity in the AG was modulated by a task where participants needed to combine concepts to form meaningful representations (Price et al., 2015). In Price et al. (2015), healthy adults needed to either combine words that formed a semantically meaningful concept (e.g. plaid and jacket) or a semantically unmeaningful concept (e.g. moss and pony). The authors found that the left AG showed increased activation for the semantically meaningful concepts, and that individual differences in the ability to combine these concepts correlated with activity in the left AG (Price et al., 2015). Adults might use this mapping and combining of concepts more efficiently than children, resulting in only finding AG activation for fully consolidated items in children (Polspoel et al., 2017), while adults already use this system for newly learned items. Future research might test this idea by changing the content of the training to arithmetic multiplications that are more heavily associated with each other, which could aid the mapping of the items with the solutions. Also, to fully investigate this idea, a longer training would be necessary to further evaluate this hypothesis.

Limitations and Future Directions

Although the current training was sufficient to show clear behavioral effects, it might not have been optimal to detect the hypothesized effects in the brain and this may be particularly so for the expected findings in the HC. Future studies with different types of training designs are therefore needed to more carefully test these hypotheses. On the one hand, it might be needed to expand the duration of the training or to postpone the post-training fMRI assessment for a few days in order to see the expected changes in brain activity in children. On the other hand, it could be that the training-related changes in the HC are brief and very time-specific. For example, Van Opstal et al. (2008) observed that during the training of ordered sequences in adults activity in the HC already changed after one scanning session. It might therefore be interesting for future studies to include intermediate scanning sessions to fully map the neural changes.

It is important to keep in mind that this study and design also has some limitations. Firstly, it is important to keep track of the sample size during the discussion of the results. Because the drop-out rate was higher than anticipated, we ended up with a smaller sample than desirable. Of the 45 participants from who we obtained informed consent, 38 completed both fMRI assessments. 12 of those 38 participants had moved too much in either or both of the fMRI assessments. Because there were two scans, the issue of too much movement played twice. In addition to the several suggestions to advance this type of study we make over the course of this discussion, we recommend further studies to include larger sample sizes.

Secondly, this study focused on children of a very narrow age range. The participating 9-10 year-olds were chosen because of their ability to retrieve single-digit items, as well as their adequate conceptual understanding of multiplication items to solve more complex items. The narrow age range also allowed us to minimize maturational and instructional effects. Using this design, we were able to contribute in how arithmetic learning in children of this age is reflected in the brain. As we only included one age group, the current data do not allow us to draw strong developmental conclusions in terms of how brain activity patterns change over longer developmental time. Future studies should therefore include different age groups to understand the development of these arithmetic learning effects in children.

Thirdly, we would briefly like to comment on the analysis of the accuracy data of the current study. To analyze these data, we applied a repeated measures ANOVA. Although this type of analysis is commonly used, these analyses should be interpreted with great caution because the assumption of normality is typically violated in accuracy data (Jaeger, 2008). Although this issue could be addressed by using (mixed) logit models, the current study had not enough data points to reliably model such a random-effects model.

Fourthly, the most dominant strategy for complex multiplication items was decomposition, for which intermediate problems are mostly solved using fact retrieval (e.g. for solving 16 x 4, 10 x 4 and 6 x 4 are easily retrieved arithmetic facts as intermediate solutions). This results in an overlap of solution strategies in two different experimental conditions (untrained multiplication items that have fact retrieval as intermediate solving strategies versus single-digit multiplication items that are solved using fact retrieval). This similarity in the two different conditions might make it difficult to find differences in the brain activity between them.

Next, it is also important to note that untrained items are more difficult and take longer to solve than the trained items. Time-on-task should be considered for the interpretations of these results for two reasons: First, previous research has found that activity in the IPS (Göbel et al., 2004) and the dorsal medial frontal cortex (Grinband et al., 2011) are sensitive to time-on-task. Both regions are part of the fronto-parietal network that shows increases in brain activity during the solution of untrained multiplication problems. Second, the regions that show increased activation for the untrained items compared to the trained items, overlap significantly with fronto-parietal regions that were found to show increased activity in conditions with higher difficulty over a range of different tasks (Fedorenko et al., 2013). This all indicates that the current findings are not specific to arithmetic and reflect more general cognitive learning mechanisms. The brain fronto-parietal brain responses to time-on-task (Göbel et al., 2004; Grinband et al., 2011) and task difficulty (Fedorenko et al., 2013) are difficult to disentangle: When one is learning something new, this will always take longer and will be more difficult. This is an inherent feature of learning novel materials, and it is hard to design an experiment that fully solves this problem.

Finally, future research might also consider to do a similar training study with an event-related design rather than a block design as we have used now. This might help, as discussed above, to further disambiguate the role of the AG in arithmetic, i.e. as fact retrieval from long-term memory, or as an circuit-breaker and attention allocator. Next to this, and in light of the limitation that untrained items will always be solved slower than trained items, an event-related design, together with more accurate response time recording in the scanner, might allow for analyses where only the activity during the solution of the items is taking into account, whether it is only a few moments to retrieve the correct answer, or whether it is the full stimulus event when there are effortful calculations taking place.

References

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience*, *9*(4), 278–291. https://doi.org/10.1038/nrn2334

Arsalidou, M., Pawliw-Levac, M., Sadeghi, M., & Pascual-Leone, J. (2018). Brain areas associated with numbers and calculations in children: Meta-analyses of fMRI studies. *Developmental Cognitive Neuroscience*, *30*, 239–250. https://doi.org/10.1016/j.dcn.2017.08.002

Arsalidou, M., & Taylor, M. J. (2011). Is 2 + 2 = 4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage*, *54*, 2382–2393. https://doi.org/10.1016/j.neuroimage.2010.10.009

Bailey, D. H., Littlefield, A., & Geary, D. C. (2012). The codevelopment of skill at and preference for use of retrieval-based processes for solving addition problems: Individual and sex differences from first to sixth grades. *Journal of Experimental Child Psychology*, *113*, 78–92. https://doi.org/10.1016/j.jecp.2012.04.014

Barrouillet, P., Mignon, M., & Thevenot, C. (2008). Strategies in subtraction problem solving in children. *Journal of Experimental Child Psychology*, *99*(4), 233–251. https://doi.org/10.1016/j.jecp.2007.12.001

Berteletti, I., & Booth, J. R. (2015). Perceiving fingers in single-digit arithmetic problems. *Frontiers in Psychology*, *6*(March), 1–10. https://doi.org/10.3389/fpsyg.2015.00226

Bloechle, J., Huber, S., Bahnmueller, J., Rennig, J., Willmes, K., Cavdaroglu, S., Moeller, K., & Klein, E. (2016). Fact learning in complex arithmetic—the role of the angular gyrus revisited. *Human Brain Mapping*, *37*(9), 3061–3079. https://doi.org/10.1002/hbm.23226

Cho, S., Metcalfe, A. W. S., Young, C. B., Ryali, S., Geary, D. C., & Menon, V. (2012). Hippocampal–prefrontal engagement and dynamic causal interactions in the maturation of children’s fact retrieval. *Journal of Cognitive Neuroscience*, *24*(9), 1849–1866.

Cho, S., Ryali, S., Geary, D. C., & Menon, V. (2011). How does a child solve 7 + 8? Decoding brain activity patterns associated with counting and retrieval strategies. *Developmental Science*, *14*(5), 989–1001. https://doi.org/10.1111/j.1467-7687.2011.01055.x

De Smedt, B., Holloway, I. D., & Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetical fluency. *NeuroImage*, *57*(3), 771–781. https://doi.org/10.1016/j.neuroimage.2010.12.037

De Smedt, B., Taylor, J., Archibald, L., & Ansari, D. (2010). How is phonological processing related to individual differences in children’s arithmetic skills? *Developmental Science*, *13*(3), 508–520. https://doi.org/10.1111/j.1467-7687.2009.00897.x

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

Delazer, M., Domahs, F., Bartha, L., Brenneis, C., Lochy, A., Trieb, T., & Benke, T. (2003). Learning complex arithmetic—an fMRI study. *Cognitive Brain Research*, *18*, 76–88. https://doi.org/10.1016/j.cogbrainres.2003.09.005

Delazer, M., Ischebeck, A., Domahs, F., Zamarian, L., Koppelstätter, F., Siedentopf, C., Kaufmann, L., Benke, T., & Felber, S. (2005). Learning by strategies and learning by drill—evidence from an fMRI study. *NeuroImage*, *25*, 838–849. https://doi.org/10.1016/j.neuroimage.2004.12.009

Eickhoff, S. B., Paus, T., Caspers, S., Grosbras, M. H., Evans, A. C., Zilles, K., & Amunts, K. (2007). Assignment of functional activations to probabilistic cytoarchitectonic areas revisited. *NeuroImage*, *36*(3), 511–521. https://doi.org/10.1016/j.neuroimage.2007.03.060

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(41), 16616–16621. https://doi.org/10.1073/pnas.1315235110

Fresnoza, S., Christova, M., Purgstaller, S., Jehna, M., Zaar, K., Hoffermann, M., Mahdy Ali, K., Körner, C., Gallasch, E., von Campe, G., & Ischebeck, A. (2020). Dissociating Arithmetic Operations in the Parietal Cortex Using 1 Hz Repetitive Transcranial Magnetic Stimulation: The Importance of Strategy Use. *Frontiers in Human Neuroscience*, *14*(July), 1–15. https://doi.org/10.3389/fnhum.2020.00271

Geary, D. C., Bow-Thomas, C. C., & Yao, Y. (1992). Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology*, *54*, 372–391.

Göbel, S. M., Johansen-Berg, H., Behrens, T., & Rushworth, M. F. S. (2004). Response-Selection-Related parietal activation during number comparison. *Journal of Cognitive Neuroscience*, *16*(9), 1536–1551.

Grabner, R. H., Ischebeck, A., Reishofer, G., Koschutnig, K., Delazer, M., Ebner, F., & Neuper, C. (2009). Fact learning in complex arithmetic and figural-spatial tasks: The role of the angular gyrus and its relation to mathematical competence. *Human Brain Mapping*, *30*(9), 2936–2952. https://doi.org/10.1002/hbm.20720

Grinband, J., Savitskaya, J., Wager, T. D., Teichert, T., Ferrera, V. P., & Hirsch, J. (2011). The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *NeuroImage*, *57*(2), 303–311. https://doi.org/10.1016/j.neuroimage.2010.12.027

Ischebeck, A., Zamarian, L., Egger, K., Schocke, M., & Delazer, M. (2007). Imaging early practice effects in arithmetic. *Human Brain Mapping Journal*, *36*, 993–1003.

https://doi.org/10.1016/j.neuroimage.2007.03.051

Ischebeck, A., Zamarian, L., Schocke, M., & Delazer, M. (2009). Flexible transfer of knowledge in mental arithmetic - An fMRI study. *NeuroImage*, *44*(3), 1103–1112. https://doi.org/10.1016/j.neuroimage.2008.10.025

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

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Lemaire, P. (2017). Cognitive Development from a Strategy Perspective. In *Cognitive Development from a Strategy Perspective*. Routledge. https://doi.org/10.4324/9781315200446

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children’s learning of multiplication. *Journal of Experimental Psychology: General*, *124*(1), 83–97. https://doi.org/10.1037/0096-3445.124.1.83

Menon, V. (2015). Arithmetic in the child and adult brain. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford Handbook of Numerical Cognition* (pp. 502–530). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.041

Menon, V., Mackenzie, K., Rivera, S. M., & Reiss, A. L. (2002). Prefrontal cortex involvement in processing incorrect arithmetic equations: Evidence from event-related fMRI. *Human Brain Mapping*, *16*(2), 119–130. https://doi.org/10.1002/hbm.10035

Nieder, A., & Dehaene, S. (2009). *Representation of Number in the Brain*. https://doi.org/10.1146/annurev.neuro.051508.135550

Peters, L., & De Smedt, B. (2018). Arithmetic in the developing brain: A review of brain imaging studies. *Developmental Cognitive Neuroscience*, *30*, 265–279. https://doi.org/10.1016/j.dcn.2017.05.002

Poldrack, R. A. (2011). Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron*, *72*(5), 692–697. https://doi.org/10.1016/j.neuron.2011.11.001

Polspoel, B., Peters, L., Vandermosten, M., & De Smedt, B. (2017). Strategy over operation: neural activation in subtraction and multiplication during fact retrieval and procedural strategy use in children. *Human Brain Mapping*, *38*(9), 4657–4670. https://doi.org/10.1002/hbm.23691

Prado, J., Mutreja, R., & Booth, J. R. (2014). Developmental dissociation in the neural responses to simple multiplication and subtraction problems. *Developmental Science*, *17*(4), 537–552. https://doi.org/10.1111/desc.12140

Prado, J., Mutreja, R., Zhang, H., Mehta, R., Desroches, A. S., Minas, J. E., & Booth, J. R. (2011). Distinct representations of subtraction and multiplication in the neural systems for numerosity and language. *Human Brain Mapping*, *32*(11), 1932–1947. https://doi.org/10.1002/hbm.21159

Price, A. R., Bonner, M. F., Peelle, J. E., & Grossman, M. (2015). Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *Journal of Neuroscience*, *35*(7), 3276–3284. https://doi.org/10.1523/JNEUROSCI.3446-14.2015

Qin, S., Cho, S., Chen, T., Rosenberg-Lee, M., Geary, D. C., & Menon, V. (2014). Hippocampal-neocortical functional reorganization underlies children's cognitive development. *Nature Neuroscience*, *17*(9), 1263–1269. https://doi.org/10.1038/nn.3788

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

Rosenberg-Lee, M. (2018). Training studies: An experimental design to advance educational neuroscience. *Neuroscience*, *Mind*, *Brain*, and *Education*, 1–11. https://doi.org/10.1111/mbe.12166

Rosenberg-Lee, M., Barth, M., & Menon, V. (2011). What difference does a year of schooling make? Maturation of brain response and connectivity between 2nd and 3rd grades during arithmetic problem solving. *NeuroImage*, *57*, 796–808. https://doi.org/10.1016/j.neuroimage.2011.05.013

Rosenberg-Lee, M., Iuculano, T., Bae, S. R., Richardson, J., Qin, S., Jolles, D. D., & Menon, V. (2018). Short-term cognitive training recapitulates hippocampal functional changes associated with one year of longitudinal skill development. *Trends in Neuroscience and Education*, *10*(June 2017), 19–29. https://doi.org/10.1016/j.tine.2017.12.001

Siegler, R. S. (1996). *Emerging minds : the process of change in children's thinking*. New York (N.Y.) : Oxford university press,.

Siegler, R. S. (2016). Continuity and Change in the Field of Cognitive Development and in the Perspectives of One Cognitive Developmentalist. *Child Development Perspectives*, *10*(2), 128–133. https://doi.org/10.1111/cdep.12173

Siegler, R. S., & Shrager, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of Cognitive Skills* (pp. 229–293). Lawrence Erlbaum.

Tschentscher, N., & Hauk, O. (2014). How are things adding up ? Neural differences between arithmetic operations are due to general problem solving strategies. *NeuroImage*, *92*, 369–380. https://doi.org/10.1016/j.neuroimage.2014.01.061

Wang, L., Liu, X., Guise, K. G., Knight, R. T., Ghajar, J., & Fan, J. (2010). Effective connectivity of the fronto-parietal network during attentional control. *Journal of Cognitive Neuroscience*, *22*(3), 543–553. https://doi.org/10.1162/jocn.2009.21210

Yang, Y., Zhong, N., Friston, K., Imamura, K., Lu, S., Li, M., Zhou, H., Wang, H., Li, K., & Hu, B. (2017). The functional architectures of addition and subtraction: Network discovery using fMRI and DCM. *Human Brain Mapping*, *38*(6), 3210–3225. https://doi.org/10.1002/hbm.23585

Zamarian, L., & Delazer, M. (2015). Arithmetic learning in adults: Evidence from brain imaging. In R. Cohen Kadosh & A. Dowker (Eds.), *The Oxford Handbook of Numerical Cognition*. Oxford University Press.