

INDIVIDUAL DIFFERENCES IN ARITHMETICAL DEVELOPMENT

EDITED BY: Ann Dowker, Bert De Smedt and Annemie Desoete
PUBLISHED IN: Frontiers in Psychology and Frontiers in Education





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-376-0

DOI 10.3389/978-2-88963-376-0

About Frontiers

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

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

INDIVIDUAL DIFFERENCES IN ARITHMETICAL DEVELOPMENT

Topic Editors:

Ann Dowker, University of Oxford, United Kingdom

Bert De Smedt, KU Leuven, Belgium

Annemie Desoete, Ghent University, Belgium

Citation: Dowker, A., De Smedt, B., Desoete, A., eds. (2020). Individual Differences in Arithmetical Development. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88963-376-0

Table of Contents

- 05 Editorial: Individual Differences in Arithmetical Development**
Ann Dowker, Bert De Smedt and Annemie Desoete
- 12 Innate or Acquired? – Disentangling Number Sense and Early Number Competencies**
Julia Siemann and Franz Petermann
- 25 Relation Between Mathematical Performance, Math Anxiety, and Affective Priming in Children With and Without Developmental Dyscalculia**
Karin Kucian, Isabelle Zuber, Juliane Kohn, Nadine Poltz, Anne Wyschkon, Günter Esser and Michael von Aster
- 38 Individual Differences in Fourth-Grade Math Achievement in Chinese and English**
Nicola A. McClung and Diana J. Arya
- 49 Mathematical (Dis)abilities Within the Opportunity-Propensity Model: The Choice of Math Test Matters**
Elke Baten and Annemie Desoete
- 65 Response-To-Intervention in Finland and the United States: Mathematics Learning Support as an Example**
Piia M. Björn, Mikko Aro, Tuire Koponen, Lynn S. Fuchs and Douglas Fuchs
- 75 Do Chinese Children With Math Difficulties Have a Deficit in Executive Functioning?**
Xiaochen Wang, George K. Georgiou, Qing Li and Athanasios Tavouktsoglou
- 86 The Association of Number and Space Under Different Tasks: Insight From a Process Perspective**
Zhijun Deng, Yinghe Chen, Meng Zhang, Yanjun Li and Xiaoshuang Zhu
- 95 Counting and Number Line Trainings in Kindergarten: Effects on Arithmetic Performance and Number Sense**
Ilona Friso-van den Bos, Evelyn H. Kroesbergen and Johannes E. H. Van Luit
- 106 Different Subcomponents of Executive Functioning Predict Different Growth Parameters in Mathematics: Evidence From a 4-Year Longitudinal Study With Chinese Children**
Wei Wei, Liyue Guo, George K. Georgiou, Athanasios Tavouktsoglou and Ciping Deng
- 116 Children's Non-symbolic and Symbolic Numerical Representations and Their Associations With Mathematical Ability**
Yanjun Li, Meng Zhang, Yinghe Chen, Zhijun Deng, Xiaoshuang Zhu and Shijia Yan
- 126 Taking Language out of the Equation: The Assessment of Basic Math Competence Without Language**
Max Greisen, Caroline Hornung, Tanja G. Baudson, Claire Muller, Romain Martin and Christine Schiltz
- 139 Individuality in the Early Number Skill Components Underlying Basic Arithmetic Skills**
Jonna B. Salminen, Tuire K. Koponen and Asko J. Tolvanen

- 150 ***Does Multi-Component Strategy Training Improve Calculation Fluency Among Poor Performing Elementary School Children?***
Tuire K. Koponen, Riikka Sorvo, Ann Dowker, Eija Räikkönen, Helena Viholainen, Mikko Aro and Tuija Aro
- 164 ***Children's Strategy Choices on Complex Subtraction Problems: Individual Differences and Developmental Changes***
Sara Caviola, Irene C. Mammarella, Massimiliano Pastore and Jo-Anne LeFevre
- 180 ***Do Exact Calculation and Computation Estimation Reflect the Same Skills? Developmental and Individual Differences Perspectives***
Dana Ganor-Stern
- 187 ***Variability in Single Digit Addition Problem-Solving Speed Over Time Identifies Typical, Delay and Deficit Math Pathways***
Robert A. Reeve, Sarah A. Gray, Brian L. Butterworth and Jacob M. Paul
- 198 ***Training Can Increase Students' Choices for Written Solution Strategies and Performance in Solving Multi-Digit Division Problems***
Marije F. Fagginger Auer, Marian Hickendorff and Cornelis M. van Putten
- 211 ***The Open Algorithm Based on Numbers (ABN) Method: An Effective Instructional Approach to Domain-Specific Precursors of Arithmetic Development***
Gamal Cerda, Estíbaliz Aragón, Carlos Pérez, José I. Navarro and Manuel Aguilar
- 223 ***Associative Cognitive Factors of Math Problems in Students Diagnosed With Developmental Dyscalculia***
Johannes Erik Harold Van Luit and Sylke Wilhelmina Maria Toll
- 232 ***Cognitive and Affective Correlates of Chinese Children's Mathematical Word Problem Solving***
Juan Zhang, Sum Kwing Cheung, Chenggang Wu and Yaxuan Meng
- 242 ***Math Anxiety in Combination With Low Visuospatial Memory Impairs Math Learning in Children***
Mojtaba Soltanlou, Christina Artemenko, Thomas Dresler, Andreas J. Fallgatter, Ann-Christine Ehlis and Hans-Christoph Nuerk
- 252 ***Effects of Working Memory, Strategy use, and Single-Step Mental Addition on Multi-Step Mental Addition in Chinese Elementary Students***
Yi Ding, Ru-De Liu, Hongyun Liu, Jia Wang, Rui Zhen and Rong-Huan Jiang
- 266 ***Heterosis in COMT Val158Met Polymorphism Contributes to Sex-Differences in Children's Math Anxiety***
Annelise Júlio-Costa, Aline Aparecida Silva Martins, Guilherme Wood, Máira Pedroso de Almeida, Marlene de Miranda, Vitor Geraldi Haase and Maria Raquel Santos Carvalho
- 283 ***Assessing Mathematical School Readiness***
Sandrine Mejias, Claire Muller and Christine Schiltz



Editorial: Individual Differences in Arithmetical Development

Ann Dowker^{1*}, Bert De Smedt^{2†} and Annemie Desoete^{3†}

¹ Experimental Psychology, University of Oxford, Oxford, United Kingdom, ² Faculty of Psychology and Educational Sciences, KU Leuven, Leuven, Belgium, ³ Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium

Keywords: mathematical cognition, individual differences, mathematical development, domain general abilities, numerical abilities, children, educational assessment, educational interventions

Editorial on the Research Topic

Individual Differences in Arithmetical Development

Individual differences in arithmetical performance have been known for a long time to be very marked in both children and adults (Dowker, 2005). For example (Cockcroft, 1982), reported that an average British class of 11-year-olds is likely to contain the equivalent of a 7-year range in arithmetical ability; and similar results were obtained 20 years and several educational changes later by Brown et al. (2002). Individual differences in arithmetic among children of the same age are also very great in most other countries. Such individual differences often appear to persist through life. At one end of the scale, about 22% of adults in the UK experience severe difficulties with basic numeracy, to an extent that leads to significant problems with employment and other everyday life activities. At the other end of the scale, some adults have an extreme fascination with numbers, can reason extremely well about numbers, and/or are exceptionally rapid and efficient calculators (Lubinski and Benbow, 2006).

There is increasing evidence that not only are there significant individual differences in children's arithmetic, but also that arithmetical ability is not unitary, but is made up of many different subcomponents (Jordan et al., 2009; Cowan et al., 2011; Desoete, 2015; Dowker, 2015; Pieters et al., 2015) and that individuals can show marked discrepancies, in both directions between different components: e.g., oral and written arithmetic; factual and procedural knowledge; exact calculation and estimation.

Individual differences in arithmetic are also increasingly studied from the point of view of their relation to more domain-general cognitive abilities, especially working memory and other executive functions. There is much evidence for significant relationships between executive functions and arithmetic (Bull and Scerif, 2001; De Smedt et al., 2009; De Weerd et al., 2013; Bull and Lee, 2014; Peng et al., 2016; Bellon et al., 2019). Most studies have looked at executive functions as predictors of arithmetic; but there is some evidence for bidirectional relationships between the two (Welsh et al., 2010; Clements et al., 2016).

Individual differences in arithmetic include not only strictly cognitive factors but emotional ones as well. (Dehaene, 1997 p. 225) pointed out that, even when studying the neural aspects of mathematics, it is important to take emotional factors into account: "...cerebral function is not confined to the cold transformation of information according to logical rules. If we are to understand how mathematics can become the subject of so much passion or hatred, we have to grant as much attention to the computations of emotion as to the syntax of reason." In particular, mathematics anxiety, sometimes amounting to real fear of mathematics is a very common phenomenon and is significantly negatively correlated with mathematical performance (Hembree, 1990; Ma and Kishor, 1997; Carey et al., 2016; Dowker et al., 2016; Foley et al., 2017; Sorvo et al., 2017; Zhang and Kong, 2019).

OPEN ACCESS

Edited by:

Peter Klaver,
University of Surrey, United Kingdom

Reviewed by:

Christine Schiltz,
University of
Luxembourg, Luxembourg
Wenke Möhring,
University of Basel, Switzerland

*Correspondence:

Ann Dowker
ann.dowker@psy.ox.ac.uk

[†]These authors have contributed
equally to this work

Specialty section:

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

Received: 23 September 2019

Accepted: 13 November 2019

Published: 03 December 2019

Citation:

Dowker A, De Smedt B and Desoete A
(2019) Editorial: Individual Differences
in Arithmetical Development.
Front. Psychol. 10:2672.
doi: 10.3389/fpsyg.2019.02672

The study of individual differences in arithmetic, from all these perspectives has important implications for mathematics education and in particular for interventions with children with mathematical difficulties (Butterworth et al., 2011; Clements and Sarama, 2011; Chodura et al., 2015; Dowker, 2017).

The articles in this special issue are extremely diverse, reflecting a very varied area; but may be divided into the following broad categories: (1) the extent, nature and persistence of individual differences in mathematics, including methods of assessing these; (2) the componential nature of arithmetical ability, and discrepancies between different aspects of arithmetical cognition and performance; (3) the relationship between arithmetic and cognitive characteristics; (4) the relationships between mathematical performance and mathematics anxiety; and (5) implications of findings about individual differences for interventions for children with arithmetical difficulties.

(1) The nature and assessment of individual differences in arithmetic.

Mejias et al. studied the assessment of early mathematical abilities in school beginners. They developed a Mathematical School Readiness test assessing early mathematical abilities. In their study, 346 children, with a mean age of 6; 3 years, were given this test entering first grade, and it was found to correlate with classical curriculum mathematics test at the time, and also to predict later performance on such tests in second grade, thus suggesting that it may be a useful test for assessing school beginners' readiness for studying mathematics, and in particular for identifying children at risk for experiencing mathematical difficulties.

Greisen et al. investigated ways of assessing mathematics that do not depend on language. This is important for children who have language difficulties, or who are receiving their instruction in a language other than their native language; and also in comparing children from countries that speak different languages. The researchers developed video and animation-based task instructions on touchscreen devices that require no verbal explanation. These tasks were administered to two groups of children in the first grade of primary school in Luxembourg. One group ($n = 96$) received verbal instructions and the other group ($n = 141$) got video instructions. One group completed the tasks with verbal instructions while another group received video instructions. Overall, the groups performed similarly, indicating that explicit verbal instructions were usually not necessary. However, there were occasions where verbal instructions were less effective than non-verbal instructions, and others where non-verbal instructions were less effective than verbal instructions.

Individual differences of course interact with age differences Caviola et al. studied children's strategy choices in solving complex subtraction problems, and investigated the effects of grade and of variations in problem complexity. Third-grade children (mean age 105.9 months) and fifth-grade children (mean age 129.8 months) solved multi-digit subtraction problems and described their solution strategies. In one experiment ($n = 155$; $n = 76$ in third grade; and $n = 79$ in fifth grade), they chose

their strategies spontaneously, and in another experiment ($n = 175$; $n = 88$ in third grade; and $n = 87$ in fifth grade), they were asked to choose between specified strategies. Fifth-grade children tended to use more efficient strategies, such as retrieval and decomposition, while third-grade children were more likely to use less efficient strategies such as counting and to rely more on the written right-to-left solution algorithm. However, all strategies were used by children in both age groups, and strategy choice was influenced by problem characteristics including problem complexity and presentation format.

Deng et al. carried out one of the few studies in this Research Topic that focussed on individual differences in adults. They investigated the Spatial Numerical Association of Response Codes (SNARC) effect in 240 adults using a parity judgment task (odd vs. even?) and a magnitude classification task (greater or smaller than 5?) for the eight numbers from 1 to 9 except for 5, which were randomly presented one at a time. Each task was carried out over 16 phases, divided into two blocks with a short interval between them, in each of which all eight items were administered. The order of the blocks was counterbalanced across participants. Detailed analyses were carried out of the changes in response times and the SNARC effect across the range of numbers and over the time course across the 16 phases. The SNARC effect emerged earlier and stayed more stable in magnitude classification task than in the parity task during the time course. It also increased over the time course in the magnitude classification task, whereas it fluctuated up and down over the time course in the parity task.

(2) The componential nature of arithmetic: how different aspects of arithmetic may diverge from one another, and how they may be influenced by different factors.

Baten and Desoete examined individual differences in primary school children's mathematics learning by combining antecedent (A), opportunity (O), and propensity (P) indicators within the Opportunity-Propensity Model (Byrnes and Miller, 2016). They studied the mathematical abilities of 114 primary school children (in grades 3–6, age range 8–12) with ($n = 61$) and without ($n = 53$) mathematical learning disabilities in relation to questionnaires given to them and to their parents and teachers. Results indicated that children with and without mathematical difficulties showed significant differences in personality, motivation, temperament, subjective well-being, self-esteem and self-perceived competence, and that there were also significant differences in parental aspirations for them. As regards antecedent (A) factors, parental aspirations explained about half of the variance in fact retrieval speed in children without mathematical learning disabilities, and socio-economic status was a strong predictor of procedural accuracy in both groups. Teachers' experience (number of years that they had taught mathematics) was considered as an Opportunity (O) factor and explained about 6% of the variance in mathematical abilities. Propensity (P) indicators explained between 52 and 69% of the variance, with intelligence as the most significant predictor overall. Indirect effects suggested that the predictors were interrelated and

highlighted the value of including A, O, and P indicators in a comprehensive model. Moreover, different A, O, and P indicators seemed to be important for fact retrieval speed compared to procedural accuracy, supporting componential theories of arithmetic.

Salminen et al. studied the early number skill profiles of 440 pre-primary Finnish children (with a mean age of 75 months), longitudinally over three points in an 8-month period. They modeled latent performance-level profile groups for three early number skill components that had been previously found to predict arithmetic (symbolic number comparison, mapping, and verbal counting skills). Four profile groups were found: lowest-performing (6%), low-performing (16%), near-average-performing (33%), and high-performing (45%). The groups differed significantly in all three number skill components and in basic arithmetic, with the lowest-performing children showing particular difficulties in the number comparison and mapping tasks, perhaps indicating problems with accessing the semantic meaning of symbolic numbers. The profiles appeared to be mostly stable over the 8-month period.

Ganor-Stern focussed in particular on the nature of exact calculation vs. computational estimation. She investigated 4th ($n = 33$), 5th ($n = 33$), and 6th grade pupils ($n = 33$) and college students ($n = 25$) performance on exact calculation and computational estimation tasks involving two-digit multiplication problems. The estimation tasks involved stating whether the result of each problem was larger or smaller than a given reference number. Older children were more accurate than younger children on the calculation task, but there were no age differences among the children for accuracy on the estimation task. There were no age differences among the children for reaction times on either task, but adults were faster than children on both. At all ages, within group variability in accuracy was greater for the exact calculation task than in the computation estimation task. Accuracy on the two tasks did not correlate strongly. The findings suggest exact calculation and computational estimation may at least in part involve different skills.

One important distinction between components of numeracy is that between symbolic and non-symbolic representations of number (Lyons et al., 2012; Schneider et al., 2017). Li et al. investigated the development of children's symbolic and non-symbolic representations of number. Participants were 253 four-to-eight-year-old children from the first and second grades of two primary schools. The researchers studied their symbolic and non-symbolic representations, their ability to map between the two types of representation, and their mathematical ability. Non-symbolic representation emerged earlier than symbolic representation, but by the age of 6, children performed equally well at both types. Children of 6 or older were able to map between symbolic and non-symbolic quantities. Path analyses showed a direct effect of children's symbolic numerical skills on mathematical performance, but non-symbolic numerical skills only affected mathematical performance indirectly via symbolic skills. The influences of symbolic and non-symbolic numerical skills on mathematical performance both decreased with age.

(3) The relationship between arithmetic and cognitive characteristics.

Wei et al. investigated the predictive role of three core executive functions (inhibition, shifting, and working memory) on the growth of mathematical skills. They carried out a 3-year longitudinal study with 179 Chinese children from second to fifth grade. In second grade with a mean age of 97.89 months, they were assessed on the above executive functions, as well as non-verbal IQ, speed of processing and number sense. Each year from second through fifth grade, they were tested on arithmetic accuracy and fluency. Structural equation modeling showed that non-verbal IQ, speed of processing, and number sense all predicted the intercept in arithmetic accuracy, while working memory was the only executive function to predict the rate of growth in arithmetic accuracy. Number sense, speed of processing, inhibition, and shifting were all significant predictors of the intercept in arithmetic fluency; but none of the executive functions predicted rate of growth in arithmetic fluency. Thus, the study suggests both that executive functions predict mathematical learning and performance, and that different executive functions may predict different aspects of mathematics.

Ding et al. studied the roles of working memory and two domain-specific factors—single-step mental addition skills, and strategy use—in multi-step mental addition in two groups of Chinese elementary students. In Study 1 ($n = 40$), they studied the effect on strategy types of task manipulations involving schema automaticity (whether intermediate sums added up to decades, e.g., convert $16 + 27$ to $16 + 24 = 40 + 3 = 43$) and working memory load (two steps vs. four steps). In Study 2 ($n = 43$), they studied the effect on strategy types of task manipulations involving schema automaticity (one-time vs. two-time regrouping) and working memory load (partial vs. complete decomposition). Results of both studies suggested that shorter response time on single-step mental addition, choice of easier strategies, and phonological working memory performance were all associated with shorter response time on multi-step mental addition. The findings in both studies highlighted the important role of the phonological loop in mental addition in Chinese children.

Siemann and Petermann discussed explanations for developmental dyscalculia, and in particular, the question of whether mathematical ability depends purely on domain-general cognitive abilities, or requires an innate number sense. They suggest that the controversy arises from ambiguity about what number sense is. They argue that it is common for early number competence to be used as a proxy for innate magnitude processing, even though it requires some knowledge of the number system (i.e., the sequence of symbols, counting words or Arabic numerals, to represent number). Thus, most studies that refer to “non-symbolic” number processing are in fact referring to tasks requiring some symbolic knowledge as well. The authors suggest that developmental dyscalculia is in fact due to a conglomerate of deficits rather than a single deficit.

Reeve et al. studied the extent to which the variability in the time children took to solve single digit addition (SDA) problems predicted their later ability to solve more complex mental addition problems; and whether children with deficits could thus be distinguished from those with typical or delayed mathematical acquisition. One hundred sixty-four children were tested on four occasions over a 6-year period starting from the age of five. They were tested on digit span, visuospatial working memory and non-verbal IQ; speed in naming single numbers and letters; speed in subitizing one to three dots; and on four occasions, speed and accuracy on a 12-item single digit addition test. At the end of the study, the children, by then aged 11, were given a double-digit mental addition test. The researchers conducted a latent profile analysis to determine if there were different variability patterns over time with regard to single digit addition. There were three distinct variability patterns. In a typical acquisition pathway, mean reaction times were relatively low and reaction time variability decreased over time. In a delayed pathway, both mean reaction time and reaction time variability started out as high, but decreased over time. In a deficit pathway, mean reaction time and reaction time variability remained high throughout the study. The deficit pathway differed significantly from the other pathways in subitizing, but not in domain-general cognitive abilities or in double-digit addition. The researchers concluded that it is important to study individual differences in reaction time variability longitudinally, and that the results highlight the importance of subitizing ability as a diagnostic index for mathematical difficulties.

Van Luit and Toll studied 84 Dutch pupils between the ages of 8 and 18, with a diagnosis of developmental dyscalculia. They looked at the prevalence in this group of deficits in four cognitive characteristics: planning skills, naming speed, short-term and/or working memory, and attention. They found that the commonest deficit was in naming speed (in particular, naming numbers), followed by deficits in short-term/working memory and planning skills. Deficits in attention were the least common.

Wang et al. investigated whether children with mathematical difficulties also experience deficits in executive functions, and whether these could be explained by lower-level deficits in processing speed. They assessed 84 children of approximately 10 years: 23 children with mathematical difficulties alone; 30 children with combined mathematical and reading difficulties; and 31 typically developing children. The children were given tests of reading, mathematics, inhibition, attentional shifting, working memory and processing speed. The children with mathematical difficulties performed worse than typically developing children on all executive function tasks. Children with only mathematical difficulties performed similarly to the children with combined mathematical and reading difficulties, except in attentional shifting, where the former performed better. However, group differences in executive functions disappeared after controlling for processing speed. Thus, it appears that most deficits in executive function, shown by Chinese children with mathematical difficulties can be accounted for by lower-level deficits in processing speed.

Mathematical ability is also considered to be influenced by language factors including both linguistic ability (Pimperton and Nation, 2010; Bjorn et al., 2016) and language background (Miura et al., 1993; Krinzinger et al., 2011; Klein et al., 2013; Dowker and Nuerk, 2016; Dowker and Li, 2019). In particular, speakers of languages with more transparent counting systems such as Chinese seem to find some aspects of mathematics easier than speakers of languages with less transparent counting systems such as English. McClung and Arya studied individual differences in 23,220 Chinese and English fourth-grade pupils mathematics achievement. They used a subset of the 2011 Progress in International Reading and Literacy Study (PIRLS) and Trends in International Mathematics and Science Study (TIMSS) data from students who were tested in Chinese or English in nine countries. Their overall scores for mathematics and reading were assessed; and their scores specifically on the Number content of the test were used to assess whether they did or did not have mathematical difficulties. Hierarchical linear modeling analyses suggested that the main effect of language on mathematical performance remained significant once their categorization as having vs. not having mathematical difficulties was added to the model. However, the effect of language on mathematical performance appeared to be especially salient in the presence of mathematical difficulties; suggesting that linguistic factors such as counting system transparency may be particularly important for children who are struggling with numeracy.

- (4) The relationships between mathematical performance and mathematics anxiety.

Kucian et al. examined the relationship between negative emotion toward mathematics and arithmetical performance in children with and without developmental dyscalculia. They studied 172 primary school children (76 with developmental dyscalculia and 96 controls). They used an affective priming task, which consisted of a simple addition or subtraction true/false decision task preceded by a prime, which consisted of words with either positive, negative, neutral affect, and words related to mathematic. It was expected that performance children with developmental dyscalculia would be slower and less accurate if preceded by a mathematics prime. In fact, neither group showed a negative mathematics priming effect, though children with dyscalculia showed lower mathematics performance than controls, and also showed more mathematics anxiety in an explicit questionnaire. Explicit mathematics anxiety correlated negatively with performance in both groups. This suggests that in primary school children, mathematics anxiety and its relation to performance may be more reliably measured by an explicit questionnaire than by a priming task. This is also suggested for university students in an unpublished study by (Dowker and Parker, 2013).

Some of the studies have looked at how the relationship between mathematical performance and mathematics anxiety may be mediated by other cognitive factors. Zhang et al. studied mathematical word-problem solving and its relation

to several cognitive and affective factors in 116 third-grade Chinese children with a mean age of 9.6 years. They found that after controlling for age and non-verbal intelligence, mathematical word problem solving correlated positively with working memory, reading comprehension and mathematical fact fluency, and negatively with mathematics anxiety. It also correlated negatively with reading anxiety, but this relationship turned out to be fully mediated by mathematics anxiety.

Soltanlou et al. studied the relationships between mathematics anxiety, visuospatial memory and mathematical learning. Twenty-five 5th graders with a mean age of 11.13 years underwent seven training sessions of multiplication over the course of 2 weeks. After the sessions, children were faster and more accurate in solving trained problems than untrained problems. Children who were both high in mathematics anxiety and low in visuospatial working memory showed worse learning than other children. This was shown specifically for accuracy, but not for reaction time. It is interesting that children with poor visuospatial working memory as well as high mathematics anxiety showed this effect. This may be because mathematics anxiety increases the load on working memory, but this only has a negative impact if working resources are already limited. We would also suggest that, as some studies have indicated (e.g., DeCaro et al., 2010), mathematics anxiety may exert its strongest effect on verbal working memory, so that visuospatial working memory may compensate for this in individuals with good visuospatial working memory, but not in those with poor visuospatial working memory.

Júlio-Costa et al. studied mathematics anxiety from a different perspective. They investigated aspects of the molecular-genetic contribution to mathematics anxiety. They looked in particular at the COMT Val158Met polymorphism, which affects dopamine levels in the prefrontal cortex, and has been found to be associated with anxiety (Hosák, 2007). Two copies of the valine allele (Val/Val) is associated with lower dopamine availability, and two copies of the methionine allele (Met/Met) with higher dopamine availability. The researchers assessed 389 school children aged 7–12 years for intelligence, numerical estimation, arithmetic achievement and mathematics anxiety and genotyped them for the COMT Val158Met polymorphism. No significant main effects were found on any of the genotype related measures. However, there were significant interactions between gender and genotype for IQ and mathematics anxiety. IQ scores were higher in Met/Met girls than in girls with at least one valine allele, though the genotype effects were not significant for boys. In the case of mathematics anxiety, heterozygous individuals tended to score close to the average, regardless of gender. Homozygous boys for either val/val or met/met showed significantly less mathematics anxiety than heterozygous boys and homozygous girls for either val/val or met/met showed significantly more mathematics anxiety than heterozygous girls.

(5) Applications of the study of individual differences in arithmetic to the development or improvement of educational practices for arithmetic teaching as a whole and/or interventions for children with difficulties.

Cerda et al. compared two teaching approaches to formal and informal mathematical reasoning with two groups of young Spanish schoolchildren ($n = 229$), aged four and five. The ABN method (Open Algorithm Based on Numbers; $n = 147$) was associated with better results than the CBC method (Closed Algorithms Based on Ciphers; $n = 82$), which is the usual approach in Spanish schools. Moreover, the effect was greater in children who received more instruction on skills considered as domain-specific predictors of later arithmetic, such as magnitude comparison and knowledge of cardinality.

Auer et al. pointed out that children have often been found to make suboptimal choices between mental and written strategies to solve division problems. In particular, lower-attaining pupils often use mental strategies where the use of written algorithms would be more efficient. They divided 147 sixth-grade pupils with low mathematics attainment into two training groups: one with explicit training to promote writing down calculations, and one which devoted a similar amount of time to practice, but without explicit targeting of strategy use. Both groups improved considerably from pretest to post-test with regard both to general performance and to selection of written strategies. However, the two training groups did not differ from one another.

Koponen et al. carried out an intervention study with elementary school children in grades 2 to 5 with poor calculation fluency (mean age: 114 months). The aim was to investigate the effects of strategy training focusing on derived fact strategies integrating factual, conceptual, and procedural arithmetic knowledge. Thus, 69 Finnish children were selected on the basis of scoring below the 20th percentile on a standardized mathematics test, and using counting-based strategies in an individual assessment. The children participated in a group based strategy training twice a week for 45 min over a 12-week period. In addition, they underwent two short weekly practice sessions for basic addition skills. Their addition fluency was assessed before and immediately after intervention, and at a 5-month post-intervention follow-up, and their progress was compared with that of two control groups: one that received a reading intervention and a business-as-usual group. The mathematics intervention group improved significantly more in addition during the intervention than either of the control groups. There was an increase in fact retrieval and derived fact strategies and a decrease in counting-based strategies in the mathematics intervention group, compared to the control groups. The effects did not, however, transfer to subtraction fluency. At 5-month follow-up the mathematics intervention group maintained their gains, but did not show further progress. They were still performing better on addition fluency than the reading intervention group, but were similar to the business-as-usual group.

Friso-van den Bos et al. divided 90 kindergarten children in the Netherlands, with a mean age of 5 years 8 months, into three groups: one trained on counting, one on number line placement, and one a business-as-usual control group. They were pre-tested and post-tested on arithmetic, counting, number lines, and number comparisons. The group trained on counting improved significantly more in arithmetic, counting, and number lines than the business-as-usual group. The group trained on

number line use did not differ significantly on any measure from the business-as-usual group.

Björn et al. investigated Response to Intervention (RTI) methods in the USA and in Finland. The authors discuss the frameworks in the two countries from the point of view of assessment and instruction. They suggest that the Finnish framework is an example of support in mathematics learning that incorporates principles of RTI, such as systematized assessment and instruction, cyclic support, and modifiable instruction. Similarly, close monitoring of student progress is also at the core of RTI in the US. Informed decision making at all levels within the system (administrative, teacher, and parental; see Fuchs and Fuchs, 2005) is provided. The basic idea of RTI in the U.S. is that the school provides the child with research-based instruction while the child is in the general education environment, and the school adjusts the intensity or nature of assessment and instruction according to the student's progress (Fuchs and Fuchs, 2005). One important difference between the American and Finnish frameworks is that the American version was primarily developed for learning difficulty identification and the Finnish version was primarily intended to re-structure the existing support services for pupils struggling with mathematics. After analyzing the similarities and differences between the American and Finnish systems, the authors conclude by discussing possibilities for further refinements of the RTI approach in both countries.

CONCLUSION

The studies in this volume support previous studies in indicating that there are marked individual differences in arithmetic at all ages from preschool to adulthood; that these appear to be related to domain-specific factors, domain-general factors and emotional factors, though there is still much controversy about how these factors interact. The studies also demonstrate that arithmetical cognition is composed of multiple components, though there may be controversy about how these are related

to one another and which components are most important; and that these findings can be put to good use in developing interventions and methods of instruction. The studies also show that findings from different countries (e.g., the UK, USA, China, and Finland) often converge to give similar results and conclusions.

Further research should expand the age groups studied, to include more work with toddlers at one end and adults at the other, and to incorporate more longitudinal studies. There should also be more work on how different components of arithmetical thinking interact with, and predict, one another and how this may change with age and instruction. There should also be further work on how domain-specific and domain-general factors interact with each other at a given time and longitudinally and the extent to which both numerical abilities and so-called domain-general abilities may be influenced by context. On the other hand, one might wonder whether the terms “domain-specific” and “domain-general” are ideal as they may sometimes be misleading. For example, it is not always what constitutes as a “domain”; so-called domain-specific predictors of one ability, such as phonological awareness being predictive for reading, are also predictive of performance in another domain, i.e., arithmetic (e.g., De Smedt et al., 2010); measures of executive function always involve the processing of certain types of stimuli (e.g., numbers), and these more specific processing differences in itself may underlie individual differences. Cultural influences on both mathematical performance and mathematics anxiety should also be explored. Finally, further progress needs to be made in the development and evaluation of interventions, and in systematically investigating whether different types of intervention may be differentially effective for children with different mathematical profiles.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

- Bellon, E., Fias, W., and De Smedt, B. (2019). More than number sense: The additional role of executive functions and metacognition in arithmetic. *J. Exp. Child Psychol.* 182, 38–60. doi: 10.1016/j.jecp.2019.01.012
- Björn, P. M., Aunola, K., and Nurmi, J.E. (2016). Primary school text comprehension predicts mathematical word problem skills in secondary school. *Educ. Psychol.* 36, 362–377. doi: 10.1080/01443410.2014.992392
- Brown, M., Askew, M., Rhodes, V., Denvir, H., Ranson, E., and Wiliam, D. (2002). “Characterizing individual and cohort progress in learning numeracy: results from the Leverhulme 5-year longitudinal study,” in *Paper Delivered at American Educational Research Association Conference* (Chicago, IL).
- Bull, R., and Lee, K. (2014). Executive functioning and mathematics achievement. *Child Dev. Perspect.* 8, 36–41. doi: 10.1111/cdep.12059
- Bull, R., and Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: inhibition, switching, and working memory. *Dev. Neuropsychol.* 19, 273–293. doi: 10.1207/S15326942DN1903_3
- Butterworth, B., Sashank, V., and Laurillard, D. (2011). Dyscalculia: from brain to education. *Science* 332, 1049–1053. doi: 10.1126/science.1201536
- Byrnes, J. P., and Miller, D. (2016). The growth of mathematics and reading skills in segregated and diverse schools: an opportunity-propensity analysis of a national database. *Contemp. Educ. Psychol.* 46, 34–51. doi: 10.1016/j.cedpsych.2016.04.002
- Carey, E., Hill, F., Devine, A., and Szucs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Front. Psychol.* 6:33. doi: 10.3389/fpsyg.2015.01987
- Chodura, S., Kuhn, J.-T., and Holling, H. (2015). Interventions for children with mathematical difficulties: a meta-analysis. *Z. Psychol.* 223, 129–144. doi: 10.1027/2151-2604/a000211
- Clements, D. H., and Sarama, J. (2011). Early childhood mathematics intervention. *Science* 333, 968–970. doi: 10.1126/science.1204537
- Clements, D. H., Sarama, J., and Germeroth, C. (2016). Learning executive function and early mathematics: directions of causal relations. *Early Child. Res. Q.* 36, 79–90. doi: 10.1016/j.jecresq.2015.12.009
- Cockcroft, W. H. (1982). *Mathematics Counts*. London: HMSO.
- Cowan, R., Donlan, C., Shepherd, D. L., Cole-Fletcher, R., Saxton, M., and Hurry, J. (2011). Basic calculation proficiency and mathematics achievement in elementary school children. *J. Educ. Psychol.* 103, 786–803. doi: 10.1037/a0024556

- De Smedt, B., Jansson, R., Bouwens, K., Verschaffel, L., Boets, B., and Ghesquiere, P. (2009). Working memory and individual differences in mathematics achievement: a longitudinal study from first grade to second grade. *J. Exp. Child Psychol.* 103, 186–201. doi: 10.1016/j.jecp.2009.01.004
- De Smedt, B., Taylor, J., Archibald, L., and Ansari, D. (2010). How is phonological processing related to individual differences in children's arithmetic skills? *Dev. Sci.* 13, 508–520. doi: 10.1111/j.1467-7687.2009.00897.x
- De Weerd, F., Desoete, A., and Roeyers, H. (2013). Working memory in children with reading and/or mathematical disabilities. *J. Learn. Disabil.* 46, 461–472. doi: 10.1177/0022219412455238
- DeCaro, M. S., Rotar, K. E., Kendra, M. S., and Beilock, S. L. (2010). Diagnosing and alleviating the impact of performance pressure on mathematical problem solving. *J. Exp. Child Psychol.* 63, 1619–1630. doi: 10.1080/17470210903474286
- Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. Cambridge: Penguin Books.
- Desoete, A. (2015). "Cognitive predictors of mathematical abilities and disabilities," in *Oxford Handbook of Mathematical Cognition*, eds R. Cohen Kadosh and A. Dowker (Oxford: Oxford University Press), 899–914. doi: 10.1093/oxfordhb/9780199642342.013.033
- Dowker, A. (2005). *Individual Differences in Arithmetic: Implications for Psychology, Neuroscience, and Education*. Hove: Psychology Press. doi: 10.4324/9780203324899
- Dowker, A. (2015). "Individual differences in arithmetical abilities: the componential nature of arithmetic," in *Oxford Handbook of Mathematical Cognition*, eds R. Cohen Kadosh and A. Dowker (Oxford: Oxford University Press), 878–894. doi: 10.1093/oxfordhb/9780199642342.013.034
- Dowker, A. (2017). "Interventions for primary school children with difficulties in mathematics," in *Advances in Child Development and Behavior*, Vol. 53: *The Development of Early Mathematics Education*, eds J. Sarama, D. Clements, C. Germeroth, and C. I Day-Hess (New York, NY: Elsevier), 255–287. doi: 10.1016/bs.acdb.2017.04.004
- Dowker, A., and Li, A. M. (2019). English and chinese children's performance on numerical tasks. *Front. Psychol.* 9:2731. doi: 10.3389/fpsyg.2018.02731
- Dowker, A., and Nuerk, H.-C. (2016). Linguistic influences on mathematics. *Front. Psychol.* 7:1035. doi: 10.3389/fpsyg.2016.01035
- Dowker, A., and Parker, S. (2013). "Mathematics attitudes, implicit and explicit anxiety and performance in an undergraduate sample," in *Poster Presented at the British Psychological Society Developmental and Cognitive Joint Sections Conference* (Reading).
- Dowker, A., Sarkar, A., and Looi, C. (2016). Mathematics anxiety: what have we learned in 60 years? *Front. Psychol.* 7:508. doi: 10.3389/fpsyg.2016.00508
- Foley, A., Herts, J.B., Borgonovi, F., Guerriero, S., Levine, S.C., and Beilock, S.L. (2017). The math anxiety-performance link: a global phenomenon. *Curr. Dir. Psychol. Sci.* 26, 52–58. doi: 10.1177/0963721416672463
- Fuchs, L.S., and Fuchs, D. (2005). Enhancing mathematical problem solving for students with disabilities. *J. Spec. Educ.* 39, 45–57. doi: 10.1177/00224669050390010501
- Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *J. Res. Math. Educ.* 21, 33–46. doi: 10.2307/749455
- Hosák, L. (2007). Role of the COMT gene Val158Met polymorphism in mental disorders: a review. *Eur. Psychiatry* 22, 276–281. doi: 10.1016/j.eurpsy.2007.02.002
- Jordan, J.A., Mulhern, G., and Wylie, J. (2009). Individual differences in trajectories of arithmetical development in typically achieving 5- to 7-year-olds. *J. Exp. Child Psychol.* 103, 455–468. doi: 10.1016/j.jecp.2009.01.011
- Klein, E., Bahnmueller, J., Mann, A., Pixner, S., Kaufmann, L., and Nuerk, H. C., et al. (2013). Language influences on numerical development—
- inversion effects on multi-digit number processing. *Front. Psychol.* 4:480. doi: 10.3389/fpsyg.2013.00480
- Krinzinger, H., Grégoire, J., Desoete, A., Kaufmann, L., Nuerk, H., and Willmes, K. (2011). Differential language effects on numerical skills in second grade. *J. Cross Cult. Psychol.* 42, 614–629. doi: 10.1177/0022022111406252
- Lubinski, D., and Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: uncovering antecedents for the development of math-science expertise. *Perspect. Psychol. Sci.* 1, 316–345. doi: 10.1111/j.1745-6916.2006.00019.x
- Lyons, I.M., Ansari, D., and Beilock, S.L. (2012). Symbolic estrangement: Evidence against a strong association between numerical symbols and the quantities they represent. *J. Exp. Psychol. Gen.* 141, 635–641. doi: 10.1037/a0027248
- Ma, X., and Kishor, N. (1997). Assessing the relationship between attitude toward mathematics and achievement in mathematics: a meta-analysis. *J. Res. Math. Educ.* 28, 26–47. doi: 10.2307/749662
- Miura, I. T., Okamoto, Y., Kim, C. C., Steere, M., and Fayol, M. (1993). First graders' cognitive representation of number and understanding of place value: cross-national comparisons – France, Japan, Korea, Sweden, and the United States. *J. Educ. Psychol.* 85, 24–30. doi: 10.1037/0022-0663.85.1.24
- Peng, P., Namkung, J., Barnes, M., and Sun, C. Y. (2016). A meta-analysis of mathematics and working memory: moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *J. Educ. Psychol.* 108, 455–473. doi: 10.1037/edu0000079
- Pieters, S., Roeyers, H., Rosseel, Y., Van Waelvelde, H., and Desoete, A. (2015). Identifying subtypes among children with developmental coordination disorder and mathematical learning disabilities, using model-based clustering. *J. Learn. Disabil.* 48, 83–95. doi: 10.1177/0022219413491288
- Pimperton, H., and Nation, K. (2010). Understanding words, understanding numbers: an exploration of the mathematical profiles of poor comprehenders. *Br. J. Educ. Psychol.* 80, 255–268. doi: 10.1348/000709909X477251
- Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372
- Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Raikkonen, E., Peura, P., et al. (2017). Math anxiety and its relationship with basic arithmetic skills among primary school children. *Br. J. Educ. Psychol.* 87, 309–327. doi: 10.1111/bjep.12151
- Welsh, J. A., Nix, R. L., Blair, C., Bierman, K. L., and Nelson, K. E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *J. Educ. Psychol.* 102, 43–53. doi: 10.1037/a0016738
- Zhang, J., Zhao, N., and Kong, Q.P. (2019). The relationship between math anxiety and math performance: a meta-analytic investigation. *Front. Psychol.* 10:1613. doi: 10.3389/fpsyg.2019.01613

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

Copyright © 2019 Dowker, De Smedt and Desoete. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Innate or Acquired? – Disentangling Number Sense and Early Number Competencies

Julia Siemann^{1*} and Franz Petermann²

¹ Department of Medical Psychology and Medical Sociology, University Medical Center Schleswig-Holstein, Kiel, Germany,

² Center for Clinical Psychology and Rehabilitation, University of Bremen, Bremen, Germany

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

Annemie Desoete,
Ghent University, Belgium
Sara Caviola,
University of Cambridge,
United Kingdom
Flávia Heloísa Santos,
Universidade Estadual Paulista Júlio
de Mesquita Filho (UNESP), Brazil

*Correspondence:

Julia Siemann
julia.siemann@uksh.de;
Julia.siemann@uni-Bremen.de

Specialty section:

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

Received: 03 August 2017

Accepted: 04 April 2018

Published: 19 April 2018

Citation:

Siemann J and Petermann F (2018)
Innate or Acquired? – Disentangling
Number Sense and Early Number
Competencies. *Front. Psychol.* 9:571.
doi: 10.3389/fpsyg.2018.00571

The clinical profile termed developmental dyscalculia (DD) is a fundamental disability affecting children already prior to arithmetic schooling, but the formal diagnosis is often only made during school years. The manifold associated deficits depend on age, education, developmental stage, and task requirements. Despite a large body of studies, the underlying mechanisms remain dubious. Conflicting findings have stimulated opposing theories, each presenting enough empirical support to remain a possible alternative. A so far unresolved question concerns the debate whether a putative innate number sense is required for successful arithmetic achievement as opposed to a pure reliance on domain-general cognitive factors. Here, we outline that the controversy arises due to ambiguous conceptualizations of the number sense. It is common practice to use early number competence as a proxy for innate magnitude processing, even though it requires knowledge of the number system. Therefore, such findings reflect the degree to which quantity is successfully transferred into symbols rather than informing about quantity representation *per se*. To solve this issue, we propose a three-factor account and incorporate it into the partly overlapping suggestions in the literature regarding the etiology of different DD profiles. The proposed view on DD is especially beneficial because it is applicable to more complex theories identifying a conglomerate of deficits as underlying cause of DD.

Keywords: dyscalculia, domain specificity, innate number sense, subtypes, early number competence

SCOPE

In the present selective review, we discuss normal and abnormal arithmetic development. We present current positions on the central questions of:

- precursors for successful mathematical education
- risk factors for low math performance
- relative contributions of domain-specific and domain-general factors
- heterogeneity of dyscalculia symptoms.

As a starting point, we will outline the current knowledge on arithmetic acquisition separately for domain-general and domain-specific contributing factors. Based on these findings, we will then explain the key deviations from the regular developmental path present in children

with dyscalculia according to the literature. For this purpose, typical findings on healthy children with regard to contributions of domain-general as well as domain-specific factors are outlined. Afterward, these are delineated from maladaptive mathematical development.

Next, we will turn to the central question of heterogeneity in developmental dyscalculia (DD). At present, there are still diverse suggested key abnormalities in the literature based on contradictory study results. From this, we will turn to an associated problem: despite the general agreement that there are subtypes of math difficulties, there is an apparent gap with respect to cognitive processes. Here, we wish to put forward that a finer distinction between innate number sense and early number competence helps in disentangling studies contradicting each other. For that purpose, we introduce a three-factor account that is based on past findings and extends previous models. We complement the above by bringing forward several potential reasons leading to different concepts of DD. Finally, we reconcile these seemingly incompatible positions by suggesting how future studies could benefit from our conception of arithmetic development and DD.

HEALTHY MATH DEVELOPMENT: INTERACTIONS BETWEEN DOMAIN-SPECIFIC AND DOMAIN-GENERAL FACTORS

Before turning to DD and its possible causes, we briefly describe how healthy math development proceeds, because theories on DD are necessarily grounded on this background knowledge. The mammalian brain seems to be equipped with an innate and preverbal ability to differentiate between quantities (e.g., Kucian and von Aster, 2015), the so-called “number sense” (Dehaene and Cohen, 1997). Humans (and other species) can learn to associate this system with symbolic number representations. The latter mechanism apparently evolves in parallel (Hyde, 2011) or hierarchically (von Aster and Shalev, 2007) into the exact and automatic recognition of small amounts of up to four or five items (“subitizing,” e.g., Henik et al., 2012; see Piazza, 2010 postulating a precursor object tracking system) and the approximate discrimination between larger quantities [“approximate number system” (ANS), e.g., Feigenson et al., 2004]. Similar theories postulate a “one system view” of number representation (Hyde, 2011). Subitizing and ANS thus refer to complementary mechanisms to differentiate small (exact) or large (approximate) numbers, i.e., distinct aspects of the number sense. In concert, they enable the comprehension of cardinality and ordinality (number concept and placement principles, Rapin, 2016). These mathematical principles are crucial for arithmetic and serve as early diagnostic markers (Gray and Reeve, 2014).

Innate basic abilities and acquired general skills both contribute to math development. Geary (2007) discriminates between so-called primary vs. secondary precursors to account for abilities we are biologically endowed with (biologically

primary) from skills shaped by environmental influences (biologically secondary). In the following, we will use the more general terms of domain-specific vs. domain-general (e.g., Karmiloff-Smith, 2015). Notably, some studies treat acquired numerical operations (e.g., calculation and arithmetic) as domain-specific (see conceptualization of Gersten and Chard, 1999), and the National Mathematics Advisory Panel even defines number sense as the understanding of the basic concept of numbers (precise representation of small and approximation of large numbers, counting skills, and simple numerical operations; National Mathematics Advisory Panel, 2008) rather than of magnitude *per se*. This example shows that skills related to early number competencies are taken as proxies for innate number abilities. To disambiguate these distinct concepts (early number competence and magnitude processing), we conceive of number sense as a pre-educational ability (following Berch, 2005) such as magnitude processing and estimation abilities. This differentiation is crucial when interpreting contradictory empirical findings and constitutes the starting point of our three-factor account. For that reason, it is important to consider both contributing factors (primary and secondary), as outlined below for healthy arithmetic development.

Domain-Specific Abilities

There are several theoretical considerations on math development. For example, von Aster and Shalev (2007) suggest a four-step-model of numerical development from discrete numerosity processing to abstract concepts of magnitude. Therein, domain-specific subitizing is a precursor of counting and subsequently for associating explicit symbolic representations (number words and Arabic digits) with the implicit number sense, culminating in the acquisition of a mental representation of numbers that is spatially organized on a mental number line. The model is based on the triple-code model of number processing (Dehaene et al., 2003) and sketches key brain structures for each developmental stage. Accordingly, there is empirical evidence for brain maturation processes during math learning with regard to structure (Zamarian et al., 2009), function (Rapin, 2016), and connectivity (Moeller et al., 2015). Yet, being explicitly formulated in the context of abnormal mathematical development, the four-step model may not cover the entire spectrum of developmental mechanisms in healthy children. More comprehensive models such as LeFevre et al.'s (2010) three-pathway model commonly schedule three precursors for math development, consisting of domain-specific quantity representation (including subitizing) and domain-general linguistic skills as well as variable indices of spatial processing (see Krajewski and Schneider, 2009; Cirino, 2011, for similar approaches).

These models incorporate the domain-specific number sense in distinct ways. Competing theories suggest either that ANS and acquired mathematical skills depend on common domain-general cognitive operations (Park and Brannon, 2014) or that their neuronal representations directly overlap (Lindskog et al., 2014), yet neither accounts for the diverse findings on the relation between ANS and math so far (see Hyde et al., 2016). This may result from the way that number sense and early number

competence are defined and especially whether ANS is assigned to one (e.g., Szűcs and Myers, 2017) or the other (e.g., Jordan et al., 2007).

While von Aster and Shalev (2007) define subitizing as an innate ability that is required for counting (i.e., number sense as we define it here), LeFevre et al.'s (2010) model treats magnitude processing as being synonymous with early numeracy knowledge. Moreover, in empirical studies, there is a tendency to collapse over these competencies (e.g., Powell and Fuchs, 2012). We believe that discrepant findings in the literature are contingent upon these different conceptualizations. Correspondingly, when operationalizing ANS as a proxy for number sense, only moderate levels of correlation with mathematical skills were found in adults (Chen and Li, 2014; Fazio et al., 2014) and in infants (Bonny and Lourenco, 2013) when measured concordantly (cross-sectional studies). Longitudinal studies further point to a genuine causal involvement, as expertise in ANS predicts later math growth (Libertus et al., 2013a). However, this relation decreases with age (Bonny and Lourenco, 2013; Fazio et al., 2014), hinting at a mediating role of the ANS. Thus, Libertus et al. (2013a) found the ANS to work indirectly via early number competencies, which then predict later math achievement. Accordingly, ANS acuity impacts on early number competence but not formal math skills (Libertus et al., 2013b), and the predictive impact of symbolic quantity measures exceeds that of non-symbolic scores (Sasanguie et al., 2012). This may also apply to evidence in the literature that math growth and ANS are apparently uncorrelated (Sasanguie et al., 2013; Szűcs et al., 2014). Accordingly, studies operationalizing domain-specific quantity processing via early number competence report a stronger correlation with later mathematical abilities (Jordan et al., 2007; Chu and Geary, 2015). Indeed, Sasanguie et al. (2015) suggest a binary magnitude system with separate modules for exact and approximate quantities. Likewise, Kucian and Kaufmann's (2009) model of number representation for healthy math development explicitly conceptualizes the increasing overlap between different quantity representations with age. The model is in line with the discussed findings as number becomes an abstract concept detached from concrete number representations. Such novel considerations are extensively discussed in a recent meta-analysis taking into account developmental shifts as well as different ANS operationalization measures (Schneider et al., 2017).

In sum, domain-specific magnitude processing (i.e., number sense) is at the heart of most contemporary models seeking to explain developmental trajectories of mathematical processing. Unfortunately, it remains a matter of debate whether magnitude processing is indeed abstract with a dedicated domain-specific module (see the discussion in Cohen Kadosh and Walsh, 2009). Novel conceptualizations of arithmetic development are in need, and existing accounts lack a comprehensive conceptualization that accounts for numerous discrepant findings in the literature (LeFevre, 2016). The matter is further complicated by the diverse influences of secondary precursors that are not easily disambiguated from potential primary causes (see Traeff et al., 2017 on this matter). Moreover, their relative contributions seem to be accompanied by an age-dependent

shift. Evidence on domain-general skills will be addressed in the following.

Domain-General Skills

While the previous paragraph stresses the importance of domain-specific precursors for healthy math development, other studies are devoted to the role of domain-general factors. Several early general skills predict later school math longitudinally, including visuospatial properties (Lauer and Lourenco, 2016; Verdine et al., 2017), intelligence (Dumontheil and Klingberg, 2012; Hornung et al., 2014), linguistic skills (Praet et al., 2013; Zhang et al., 2014), executive control (Bull et al., 2008; Clark et al., 2013), and working memory (LeFevre et al., 2013; Bailey et al., 2014; but see Fuchs et al., 2006). While working memory span has often been considered essential to math skill levels, this seems to be content-specific. In fact, visuospatial rather than verbal WM skills correlate with math achievement in healthy populations (Clearman et al., 2017), whereas patients with DD show stronger correlations with verbal WM (Mammarella et al., 2013). Accordingly, Szűcs (2016) identified type of WM impairment (verbal and visuospatial) as contributing to the specific profile of mathematic problems in DD patients. Moreover, the correlation between WM and math may be stronger in children with low number sense capabilities than healthy controls. Therefore, differentiation between control groups and children with DD is essential when examining domain-general factors. Thus, Szűcs et al. (2014) found no correlation between WM and math performance in healthy children. A possible explanation is given by the development of an arithmetic fact memory. Healthy individuals may be able to use their number sense to develop early number competencies (i.e., connections between magnitude and numbers, basic arithmetic principles, etc.) as a basis for an arithmetic fact memory. By contrast, children with DD cannot profit from such automated processes, rather relying on immature mental calculation strategies such as counting. These in turn draw heavily on verbal WM capacities (Alloway et al., 2006), probably leading to a stronger connection between arithmetic and WM. Correspondingly, WM seems to be especially important for more sophisticated math operations such as subtraction (Caviola et al., 2014). Finger counting may serve as a compensatory function to offload WM (Crollen et al., 2011) and is frequently observed in DD (Attout and Majerus, 2015). Nonetheless, domain-specific abilities still contribute to later math outcomes over and above general cognitive influences. Thus, elementary and middle school addition both correlate with early number comparison skills irrespective of working memory, visuospatial skills, linguistic performance, and IQ (Bailey et al., 2014). Moreover, early enumeration capacity uniquely accounts for arithmetic achievement when controlling for working memory and executive functions (Gray and Reeve, 2014). A recent meta-analysis further suggests that early number competence but not WM predicts calculation performance in at-risk children (Peng et al., 2016a). In addition, while math training programs were found to have the largest effects on early number competence, improving domain-general cognitive skills does not seem to transfer to enhanced mathematical achievement (see Raghubar and Barnes, 2017).

Emphasizing the developmental nature of healthy mathematical acquisition may help to reconcile these findings, as both factors (domain-specific and domain-general) apparently contribute to distinct aspects during math growth (Hornung et al., 2014). As a synopsis of the presented considerations, there are many influential factors on arithmetic acquisition: age of participants may determine whether domain-specific or domain-general performance predominantly correlates with math achievement; test format (verbal and visual) may lead to different results especially with respect to correlations with WM; and sample type (healthy and DD) seems to lead to different correlations due to distinct strategies. With this knowledge in mind, the following paragraphs will point to domain-specific and domain-general abnormalities during mathematical development that may cause the disorder labeled DD.

DEVELOPMENTAL DYSCALCULIA

Nomenclature

So far, there is no unitary expression for DD, as it is a complex disorder that may be associated with diverse problems including low math performance, low counting skills, weak arithmetic, struggles with calculation, or inability in understanding mathematical procedures. Accordingly, synonyms such as “persistent mathematical difficulties” (Morgan et al., 2016), “mathematics learning disability” (Murphy et al., 2007), or “mathematical difficulties” (Schwenk et al., 2017) may be used to delineate profound (maybe innate) magnitude processing from presumably acquired problem with arithmetic (see Morgan et al., 2016). This diversity of expressions for the same basic collection of symptom already indicates that there is neither a unified concept of mathematical disorders nor a consistent etiological explanation thereof. The term “mathematics learning disability” stresses the role of domain-general operations in learning mathematical proficiency and is predominantly applied by opponents of an innate number sense problem (e.g., Rousselle and Noël, 2007). By contrast, “mathematical difficulties” seem to represent a severity-based expression, leaving open the possibility for an innate as well as an acquired etiology.

The current version of the ICD-10 [International Classification of Diseases; World Health Organization [WHO], 1992] classifies dyscalculia among the pervasive and specific developmental disorders (chapter F8) as a specific developmental disorder of scholastic skills (sub-chapter F81) as a mathematical disorder (F81.2) with no further specification. The criteria demand a discrepancy between a child’s intelligence level and a standardized math test score as well as adequate mathematical educational circumstances. By contrast, in the latest version of the Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-5; American Psychiatric Association, 2013), DD is listed as *specific learning disorder with impairment in mathematics* (315.1) and may be grounded on problems with the number sense as well as with arithmetic fact retrieval, calculation, or math reasoning. The release of the revised version ICD-11 is still pending. It is to be expected that similar alterations with respect to intelligence level and mathematical scores will be put forward.

In the following, we will summarize both well-established and more recent positions on pathological math development separately for potential domain-specific and domain-general precursor abilities and integrate the gathered knowledge into a more precise view on DD.

Presumable “Domain-Specific” Symptoms Associated With DD

Developmental dyscalculia in children is characterized by profound difficulties with various fields of mathematics, including counting principles, transcoding between number digits and number words, comprehension of number syntax, numerical fact knowledge, and fact retrieval (Jordan et al., 2003). A deficient *number sense* is most frequently related to DD (according to Mazzocco and Thompson, 2005), and correlations between ANS acuity and math proficiency apparently exist prior to mathematical education (Mazzocco et al., 2011; Libertus et al., 2013a) and ANS has a predictive role for math performance in young children (Wong et al., 2017). Analogous to evidence for healthy math development (e.g., Bailey et al., 2014), numerical competence (commonly considered to be domain-specific) of at-risk children uniquely predicts math performance during elementary school even when controlling for domain-general skills (Peng et al., 2016b). By contrast, in adults, DD primarily reflects domain-general fact retrieval deficits and weak phonological processing via an impaired association between both (De Smedt and Boets, 2010).

Immature *counting* strategies in DD may be causally related to deficient fact knowledge by hindering the build-up of associations between arithmetic operations and solutions (Geary et al., 2012). Similar findings highlight the importance of progressing from procedure-based counting to memory-based *fact retrieval* (De Visscher and Noël, 2016). Thus, patients with DD are hypersensitive to interference from neighbor problems in multiplication, posing an indirect negative effect by preventing the successful storage of symbol-response associations in long-term memory (De Visscher and Noël, 2016). Still, while children with DD are particularly impaired in grasping the principle of cardinality during counting (Rapin, 2016), which is also predictive of counting in healthy arithmetic development (Moore et al., 2016), the same skill seems to play only a minor role in healthy children. They likewise fail to comprehend this principle despite otherwise healthy mathematical development (Kamawar et al., 2010). Kuhn et al. (2016) infer that DD essentially reflects a deficit of *specific precursor* abilities that healthy infants are endowed with even before learning arithmetic or calculation (i.e., estimation, enumeration, and transcoding). Longitudinal studies accordingly show that basic quantity-based abilities including number naming, counting, and estimation are stable predictors of arithmetic proficiency during the transition from preschool to kindergarten (VanDerHeyden et al., 2006), elementary school (Methe et al., 2008; Lembke and Foegen, 2009), and high school (Siegler et al., 2012), independent of general intelligence levels (Locuniak and Jordan, 2008). It is therefore likely that quantity processing enables more sophisticated mathematical manipulations. However, a recent meta-analysis

found symbolic rather than non-symbolic quantity processing measures to be related to low mathematical performance in children with DD (Schwenk et al., 2017). This is in line with the relative contributions of the number sense and of early number competence for healthy arithmetic skills discussed above. The shift from non-symbolic ANS to basic symbolic skills observed in healthy children seems to be aberrant in DD. It is therefore essential to consider supportive skills in contributing to abnormal mathematical development as well. To meet this demand, the following section is devoted to domain-general skills in the context of low mathematical abilities.

Domain-General Deficits in DD

Apart from poor numerical abilities, low math performance may also be grounded on malfunctioning supportive cognitive operations. Longitudinal and cross-sectional studies repeatedly identified performance differences between children with DD and healthy controls in attention (Ashkenazi et al., 2009; Swanson, 2011), executive functions (especially inhibitory control; Swanson, 2012; Szűcs et al., 2013), linguistic skills (as mediator, see Jordan et al., 2015), intelligence (Geary et al., 2008; but see Alloway, 2009), general processing speed (Geary, 2010; Namkung and Fuchs, 2016), and visuospatial processing (Hanich et al., 2001), especially with respect to short-term (Andersson, 2010) or working memory (Ashkenazi et al., 2013; Bugden and Ansari, 2015). Yet evidence on their individual relative contributions is inconclusive, and only few studies systematically explored single cognitive factors in the context of DD (Morgan et al., 2016).

As described initially, diverse skills such as language (Zhang et al., 2014), attention (Morgan et al., 2016), and intelligence (Geary and Moore, 2016) contribute to sophisticated (i.e., healthy) math knowledge. However, findings on healthy mathematical development have limited values for the etiology of DD. Especially with regard to procedural knowledge, evidence on DD is still in need. To our knowledge, only three studies have specifically addressed calculation development (reflecting procedural knowledge) in DD in comparison with healthy controls using longitudinal data so far (according to Peng et al., 2016b). Of these, Alloway (2009) found a correlation between working memory (but not intelligence) and calculation, but without controlling for other variables such as verbal skills or executive functions. In a broader approach, Namkung and Fuchs (2016) found processing speed and attention to predict later calculation expertise in DD, whereas language skills did not explain additional variance. The only study we identified (Peng et al., 2016b) that addressed this matter with data from elementary school (i.e., early math development) suggests that

- (a) processing speed (domain-general) and early number competence predict mathematical growth in DD
- (b) early number competence mediates the degree to which DD persists at the end of elementary school
- (c) children with DD and comorbid reading disorder compensate their deficits using early number competence rather than domain-general skills.

Peng et al. (2016b) also report independence between linguistic skills and whole number math in DD. In contrast to this finding, others suggest that linguistic deficits may be linked with DD in school-aged children (Fuchs et al., 2006), paralleling the relation between language and healthy math development. Consequently, assuming a direct link between reading disorder and DD – potentially with a causal relationship – is tempting. Yet evidence in this field is contradictory, as longitudinal positive correlations between both clinical samples (Jordan et al., 2002) stand against contrary findings (Andersson, 2010). Moreover, a potentially underlying impact of linguistic deficits on arithmetic fact retrieval in DD (according to Simmons and Singleton, 2008) could not be substantiated unequivocally (e.g., Geary et al., 2012).

This list of deficits associated with DD is far from complete and demonstrates how intricate it is to interpret low math performance in the broader context of mathematical development, especially when compared to healthy children and adult mathematics. This discrepancy stresses the importance of a multifactorial approach in the etiology of DD. In the next sections, we will outline opposing viewpoints in the literature with regard to the question whether abnormal mathematical development is necessarily caused by an underdeveloped number sense and will show how a finer distinction between different precursors for distinct DD subtypes can reconcile ambiguous study results.

Heterogeneity in DD

In light of the above findings, it appears that the precursors for successful math development differ from those abilities frequently impaired in DD. Healthy math skills are likely to be continuously distributed, whereas DD constitutes profound deficits distinct from the low end of this continuum (see Desoete et al., 2012). Indeed, whereas healthy math performance scores are highly variable from preschool to elementary school (Geary et al., 2000) and persistent interindividual differences only emerge at grade 2 (Jordan et al., 2003), possibly based on changing strategies (Aunola et al., 2013), DD is stable over time (Andersson, 2010). The most obvious demarcation between healthy math development and DD is evident when considering that low initial numerical competence in elementary school is often not clinically significant in follow-up tests anymore (Desoete et al., 2012). Obviously, manifold reasons can account for weak performance in math tests and need to be identified before erroneously diagnosing DD (see Kaufmann et al., 2013 for a discussion). Nonetheless, growth of mathematical proficiency depends decisively on an individual's initial numerical competence even before school (Jordan et al., 2009), suggesting that mathematical cognition may be less unitary than conceptualized in many studies (see Dowker, 2008 on individual differences).

Unfortunately, research so far lacks insight into early developmental influences of deficient precursors specifically in DD, because most studies either address later developmental stages (elementary school) or apply cross-sectional study designs impeding a proper analysis of causal influences (see Peng et al., 2016b). However, children's age represents a major contributing factor to the causes of DD and correlations with other skills.

Thus, certain precursors are only transiently related to DD (Kniesel et al., 2011), and despite the fundamental role of the so-called number sense, domain-general influences must be taken into account to differentiate between DD profiles (Szűcs, 2016). For example, domain-general visuospatial (Passolunghi et al., 2008) and decoding skills (Peng et al., 2016b) contribute to arithmetic acquisition but not later math proficiency. Likewise, the predictive role of the ANS decreases with age (Szkudlarek and Brannon, 2017), and while early number competency emerge as an initial predictor, domain-general skills gain more importance through arithmetic development (Geary et al., 2017). Failing to control for such transitory effects may in turn result in contradicting findings such that children with DD can demonstrate age-adequate domain-specific number processing competence. Accordingly, children diagnosed with DD in grade 2 showed comparable number processing profiles compared with a control group in grade 4 (Landerl and Kölle, 2009). Data from Fuchs et al. (2010) suggest that it is crucial to be cautious about the manifest variables chosen to operationalize DD. In that study, only mathematical *word problem* skills varied with basic numerical abilities, whereas *calculation* performance did not correlate with other domain-general or domain-specific variables (Fuchs et al., 2010). Neuroimaging findings suggest that the bilateral inferior parietal lobule executes domain-specific magnitude processing (Dehaene et al., 2003) and exhibits disparities in DD (Mórocz et al., 2012). However, the same structure is also engaged in domain-general skills that contribute to arithmetic, like working memory (Dumontheil and Klingberg, 2012), attention (Vandenberghe et al., 2012), and spatial processing (Yang et al., 2012). This emphasizes the diversity of DD profiles and leads to an important question raised in the literature about the etiology of distinct DD subtypes (see Andersson and Östergren, 2012 for a review).

Etiology and Subtypes

In the last decades of DD research, four distinct classes of theories have emerged (according to Castro-Canizares et al., 2009). The first suggests that a domain-specific number sense deficit underlies DD, either for approximate and analogous quantities (number sense deficit, Wilson and Dehaene, 2007) or for exact and discrete representations thereof (defective number module, Butterworth, 2005a).

Alternatively, DD may stem from poor access to quantity information, i.e., an aberrant communication between brain regions devoted to magnitude and its symbolic representation (access deficit, Rousselle and Noël, 2007).

The third class proposes a generalized magnitude system in the brain (comprising both exact and abstract quantities and extending to numbers, time, and space) that is malfunctioning in persons with DD (a theory of magnitude, Cohen Kadosh et al., 2008).

Finally, a fourth class of theories identifies a causal relation between mal-efficient domain-general factors and DD symptoms (cognitive deficits, Geary et al., 2007).

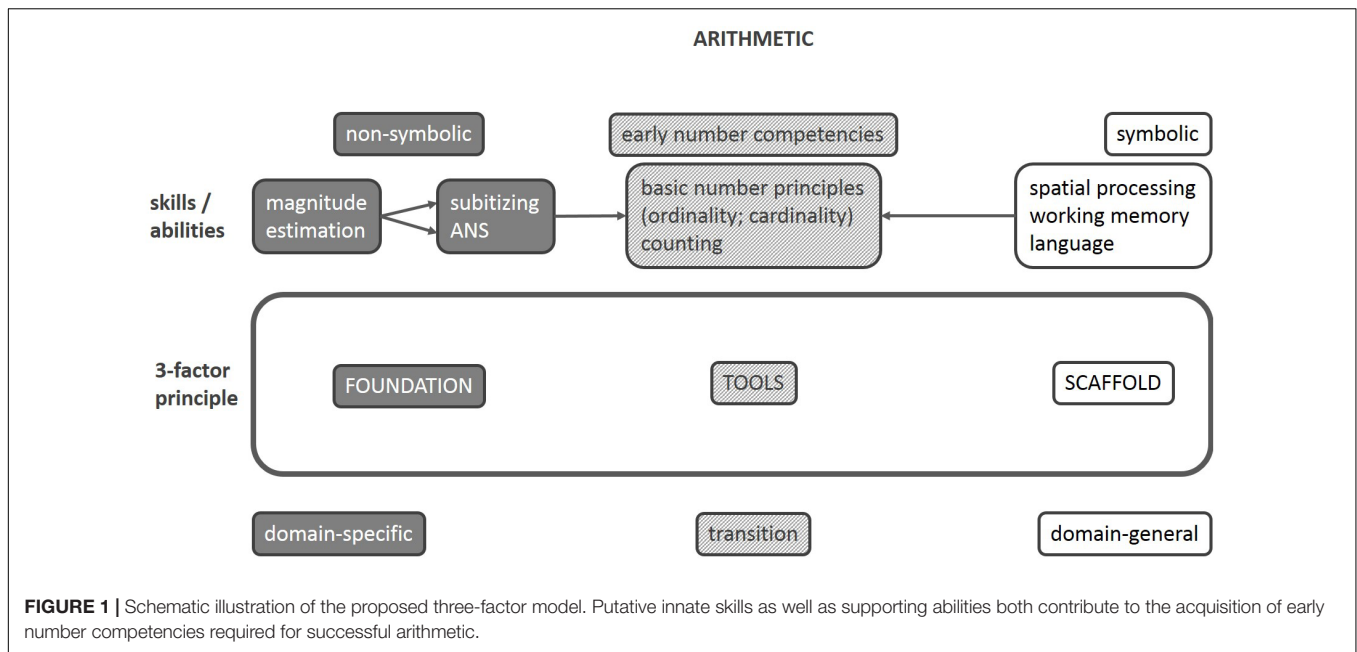
At the interface of these accounts, double deficit theories assume that deficits of multiple neuropsychological abilities contribute to learning disabilities in general

(Wolf and Bowers, 1999). However, there are no consistent findings in the literature. Accordingly, whereas rapid automatized naming of digits and phonological awareness did not predict DD in a previous study (Heikkilä et al., 2016), another study found similar operations (processing speed and verbal comprehension) to correlate with DD symptoms (Willcutt et al., 2013). Moreover, low performance in number comparison tasks is inconclusive with regard to the underlying deficit, because while this problem may stem from a defective innate number processing system (Butterworth, 2005b), an alternative explanation is a deficit in accessing this module (Rousselle and Noël, 2007). Instead, distance and problem size effects may be more informative, as they typically alleviate with development (Holloway and Ansari, 2008) and may be underdeveloped in DD (Skagerlund and Traff, 2016). In addition, there are hardly any physiological findings to support the domain-specific theory on DD (Szűcs, 2016). In fact, there is evidence that both domain-general skills and domain-specific abilities represent superordinate predictors of DD (Toll et al., 2016) that are sensitive to training programs (Kuhn and Holling, 2014). Possible reasons may be a potential dependency of number sense performance on WM during early arithmetic development (Vandervort, 2017) or the fact that tests of ANS (representing the number sense) often cannot disentangle perceptual factors drawing on WM from actual numerical skills (Bugden and Ansari, 2015).

The multitude of DD profiles may actually be grounded on separate (and potentially overlapping) etiologies (e.g., Kucian and von Aster, 2015; Skagerlund and Traff, 2016) reflected at first sight in common deficits in arithmetic performance. Thus, the above classes of hypotheses potentially apply to distinct DD phenotypes and consequently to different underlying causes: whereas a “defective module” (Butterworth, 2005a) or deficient “number sense” (Wilson and Dehaene, 2007) implies that abnormal mathematical development results from an immature magnitude representation; the latter is intact according to the “access deficit” theory (Noël and Rousselle, 2011), which centers on problems retrieving numerosity from symbolic representations (Rousselle and Noël, 2007). Therefore, distinct theories can co-exist and need not be mutually exclusive when more closely investigating the underlying deficits and their operationalization. In the following, we will show that a finer separation of domain-specific deficits dissolves several related issues in DD research.

A Novel Concept of DD Typology

We suggest that properly characterizing arithmetic development and DD require three factors – as opposed to two in the literature (see **Figure 1**). Factors 1 and 2 have previously been described. The domain-specific number sense (F1) likely represents the *foundation* on which arithmetic development rests. During formal math education, various domain-general skills (F2) assist in linking abstract numerosity with symbolic number representations, analogous to a *scaffold*. The resultant early number competence (F3) comprises *tools* that are involved in arithmetic operations.



DD may be caused by different underlying deficits: analogous to a house construction, either the foundation itself is underdeveloped (F1), making it necessary to resort to domain-general skills as a scaffold (F2), which is less stable when lacking a proper foundation. But in other cases of DD, there are low-level domain-general skills such as working memory: despite an even foundation, an instable scaffold leads to an unsteady house. Third, if the link between non-symbolic and symbolic representations of number (F3) cannot be established, this is analogous to a craftsman using broken tools.

The empirical findings we outlined above are transferable to our three-factor account. Some children with DD show deficits hinting at a poor foundation (magnitude processing; F1), whereas other DD profiles rather accord with low scaffolding support (e.g., working memory or processing speed; F2). While this distinction is well established (Andersson and Östergren, 2012), our view on arithmetic development provides a novel approach to different DD patterns because it highlights the fact that math deficits may be present despite healthy domain-general and domain-specific skills. Crucially, early number competence (F3) is often subsumed under what we consider to be domain-specific skills (F1; as outlined above). By dissociating these two qualitatively different forms of cognition, future studies may succeed in disambiguating DD subtypes. In addition, contradictions between past studies are likely the cause of misconceptions of what a number sense is and what it is not, and distinct correlation patterns between ANS and formal (fact knowledge) compared with informal math (counting) yield empirical support showing a high face validity of this proposed concept (Libertus et al., 2013b). Thus, following Kolkman et al. (2013), domain-specific number sense (F1) primarily assists in establishing a successful mapping between magnitude and symbolic representations, i.e., prior to and during early arithmetic acquisition, for which there is empirical support (Inglis et al., 2011). These considerations are

in line with the developmental model of number representation (Kucian and Kaufmann, 2009) introduced previously: while this descriptive model poses a solution for developmental changes observed with increasing arithmetic education, our three-factor account delivers a causal explanation for discrepant findings not only longitudinally (i.e., between different age groups) but also between conceptually different study designs within age groups. That transition appears to rely on domain-general skills (according to Namkung and Fuchs, 2016). Similarly, Hornung et al. (2014) postulate that early number competence in infants results from interacting basic quantity skills with domain-general abilities. We suggest that DD theories centered on domain-general deficits are primarily applicable to arithmetic acquisition and may therefore be considered valid especially in accounting for the high variability between age groups both in healthy and in clinical samples (see the discussion in Kaufmann et al., 2013).

Evidence for a third influencing factor (F3) comes from studies hinting at maleficent white matter tracts associated with poor math skills. Both interhemispheric fiber tracts between the IPL (representing the number sense, Cantlon et al., 2011) and intrahemispheric associations between IPL and angular gyrus could be verified (Klein et al., 2013).

Furthermore, the three-factor idea helps reconciling extreme positions of magnitude-based arithmetic vs. direct symbolic activation. The latter is assumed in the “encoding-complex model” (Campbell and Clark, 1988), which neglects domain-specific magnitude representations due to direct activation of numerosity based on parallel relative contributions of number representations. Thus, the foundation of early numerosity (F1) may seemingly become obsolete because studies often test early number competence (number acuity) with symbolic representations (F3). Indeed, recent findings hint at bidirectional correlations between number acuity (F3) and math skills (Lyons

et al., 2012). While at first sight, this seems to contradict the assumed innateness of number sense processing (F1), it indeed strengthens our stance of a third independent factor of number acuity that was conceptualized as number sense in Lyons et al. (2012).

In addition, our account is sensitive to developmental shifts and therefore provides a high degree of flexibility. Future research should more clearly distinguish between the different concepts in order to differentiate between DD that is based on deficits with early numerosity from innate magnitude problems. Thus, Kaufmann et al. (2013) coined the expression *secondary DD* for low mathematical skills determined or even caused by low non-numerical cognitive skills. The matter is further complicated by findings that ANS (as a proxy for magnitude) and school-based math are reciprocally related (e.g., Nys et al., 2013; but see Zebian and Ansari, 2012) and that mathematical education impacts on ANS (Piazza et al., 2013; Lindskog et al., 2014). While such results question the assumed innateness of magnitude processing, they likewise provide implications for interventions. If numerosity (in terms of early number competence) turns out to be trainable, low-performing infants should be identified already before formal schooling and participate in specific training programs. Thus far, positive outcomes of such trainings on number acuity (e.g., Nys et al., 2013; Park and Brannon, 2013) could not be established beyond doubt (e.g., Zebian and Ansari, 2012; Obersteiner et al., 2013; see Szűcs and Myers, 2017 for a review).

LIMITATIONS

The articles on DD discussed in the present review are heterogeneous with respect to many aspects impacting on the study results. In the following, we will briefly outline the associated approaches and identify potential strengths and weaknesses:

Design Considerations

Group Contrasts (Children With/Without DD)

These studies contrast children with DD and healthy age-matched controls with respect to various variables of interest (e.g., mathematical precursors; working memory; and language) and investigate the variance between both samples that each explains. While most efficient in terms of temporal and economic matters, such analyses provide little transferable information (small samples) and no basis to characterize putative causal relationships about developmental trajectories. For these purposes, cross-sectional and longitudinal studies are the means of choice.

Cross-Sectional vs. Longitudinal Studies

Both study types serve to reveal potential developmental processes. Cross-sectional studies offer a time-economic way to compare different developmental stages with each other but do not enable predictions about causal relations between possible precursor skills and later math achievements, in contrast to longitudinal studies.

Methodological Considerations

Choice of Independent Variables

Another important issue concerns the choice of cognitive functions representing putative precursors and supportive skills for successful arithmetic development. Studies often either investigate domain-specific abilities (Chen and Li, 2014; Fazio et al., 2014) or domain-general skills (LeFevre et al., 2013; Bailey et al., 2014), even though controlling for one factor in the context of another or directly contrasting both (multidimensional approach, see Szűcs, 2016) may deliver a more comprehensive insight (Aunola et al., 2004). In addition, studies concentrating on one category (i.e., specific or general) often fail to take into account a sufficient amount of associated factors.

Choice of Dependent Variables

There is no unitary operationalization of math proficiency or achievement, nor are age and school-based development adequately accounted for. As a result, heterogeneous constructs such as number system knowledge (Sowinski et al., 2015), timed math (Sasanguie et al., 2013), mental calculation (Reeve et al., 2015), standardized math test scores (Chang et al., 2015), or arithmetic fluency (LeFevre et al., 2013) exist for the same overall latent variable termed math proficiency.

Considerations on Sample Criteria

Developmental Trajectories

Irrespective of study design, the classification and comparability of participant sub-groups impact on the associated gain of knowledge. In arithmetic research, longitudinal studies often attend healthy children during the transition from kindergarten to preschool or primary school, thus allowing predictions about regular mathematical proficiency (e.g., VanDerHeyden et al., 2006; Lembke and Foegen, 2009; Siegler et al., 2012). However, the informational value in terms of developmental trajectories of mathematical disorders is limited. Consequently, longitudinal studies on children with low initial number processing abilities (number sense) and potential struggles arising with symbolic representations thereof (transcoding) are more suitable. For this, screening instruments are required that test preschoolers on non-symbolic number processing (e.g., mental number line or non-symbolic quantity estimation tests).

DD Definitions and Diagnostic (Cut-Off) Criteria

As with math proficiency, the criteria required to sort children into DD (sub)groups are equally inconsistent (see Murphy et al., 2007 for a review). Some studies use the term “persistent mathematical difficulties” to dissociate putative genuine DD from mild and potentially transitional numerical difficulties (Morgan et al., 2016), whereas others collapse over these categories (e.g., Murphy et al., 2007). Furthermore, studies seldom take into account ICD-10 diagnostic criteria for DD, and the associated cut-off criteria are commonly weakened, ranging from the 10th to the 35th percentile (Mazzocco and Thompson, 2005). Thus, qualitative differences between persistent and transient arithmetic weaknesses (Mazzocco and Myers, 2003) impede the comparability between study samples that are based on moderate (e.g., Jordan et al., 2003; Geary, 2004) or low math achievement

scores (Mazzocco et al., 2011). Besides, reporting comorbid deficits is no established practice, even though these pose additional and fundamental developmental challenges (Szűcs, 2016). In addition, concerns have been raised in the past due to the conceptual overlap between mathematical tests with IQ subtests. As a recent advance, the discrepancy criterion was abolished in the United States in 2013 (Schulte-Körne, 2014). DD research mainly relies on convenience samples (often school samples) where standardized IQ indices are not reported at all, or otherwise intelligence level was included as a domain-general regressor. In this respect, weakening the diagnostic criteria as frequently done in DD studies may be advantageous. However, lowering the required discrepancy between participants' math test score and their age-based reference group may be more problematic. Assuming that math skills are dimensionally distributed, this approach may falsely sort healthy low performers into the group of DD patients. This matter is further complicated by the ambiguous definition of DD and potential subgroups.

Considerations on Selected Studies

The present article should not be misunderstood as a systematic or exhaustive review nor does it make the claim to cover all relevant open questions about DD. The choice of studies is selective and we may not have covered all relevant viewpoints or theories on the related issues. Rather, we wish to point to one essential gap in the approach to research in this field. By drawing attention to the ambiguous conceptualization of “the number sense,” we hope to initiate a finer distinction between the discussed abilities/skills in shaping arithmetic sophistication. This may provide new ways of interpreting study results and help reconciling discrepant findings.

In sum, the available studies on DD and math development are confounded with many influential factors. This article served to sensitize researchers in this matter by contrasting evidence and standpoints in the literature from many angles.

REFERENCES

- Alloway, T. P. (2009). Working memory, but not IQ, predicts subsequent learning in children with learning difficulties. *Eur. J. Psychol. Assess.* 25, 92–98. doi: 10.1027/1015-5759.25.2.92
- Alloway, T. P., Gathercole, S. E., and Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Dev.* 77, 1698–1716. doi: 10.1111/j.1467-8624.2006.00968.x
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edn. Washington, DC: APA. doi: 10.1176/appi.books.9780890425596
- Andersson, U. (2010). Skill development in different components of arithmetic and basic cognitive functions: Findings from a 3-year longitudinal study of children with different types of learning difficulties. *J. Educ. Psychol.* 102, 115–134. doi: 10.1037/a0016838
- Andersson, U., and Östergren, R. (2012). Number magnitude processing and basic cognitive functions in children with mathematical learning disabilities. *Learn. Individ. Differ.* 22, 701–714. doi: 10.1016/j.lindif.2012.05.004
- Ashkenazi, S., Black, J. M., Abrams, D. A., Hoeft, F., and Menon, V. (2013). Neurobiological underpinnings of math and reading learning disabilities. *J. Learn. Disabil.* 46, 549–569. doi: 10.1177/0022219413483174
- Ashkenazi, S., Rubinsten, O., and Henik, A. (2009). Attention, automaticity, and developmental dyscalculia. *Neuropsychology* 23, 535–540. doi: 10.1037/a0015347

We complemented these considerations by introducing a novel approach that equally applies to the interpretation of contradictory study results as to the classification of DD subtypes. Thereby, we wished to close this gap and answer some of the questions that follow when looking at individual study results.

OUTLOOK

In order to differentiate between genuine DD and low math abilities, individual developmental trajectories should be considered in the context of various contributing skills. This idea is pressing given the broad field of domain-general and domain-specific precursors that each demonstrates interindividual differences. Disentangling low but healthy math performance from clinically relevant and persistent DD is essential and requires multilevel diagnostic instruments. These in turn depend on the identification of unique precursors of DD that should be screened early on in preschool. For that purpose, future studies are needed that address math development prior to formal mathematical education. As for now, the majority of studies examined school-aged samples, i.e., after having acquired the basic concepts of arithmetic. So far, findings on early number competence (before kindergarten) are still lacking (Morgan et al., 2016). Such studies would help to further disentangle innate abilities (F1) from acquired numerical skills (F3). Furthermore, contradictions between existing studies can possibly be reconciled in a meta-analysis when introducing our three-factor approach.

AUTHOR CONTRIBUTIONS

JS wrote the preliminary draft. JS and FP revised the manuscript and read and approved the final manuscript.

- Attout, L., and Majerus, S. (2015). Working memory deficits in developmental dyscalculia: the importance of serial order. *Child Neuropsychol.* 21, 432–450. doi: 10.1080/09297049.2014.922170
- Aunola, K., Leskinen, E., Lerkkanen, M., and Nurmi, J. (2004). Developmental dynamics of math performance from preschool to grade 2. *J. Educ. Psychol.* 96, 699–713. doi: 10.1037/0022-0663.96.4.699
- Aunola, K., Tolvanen, A., Viljaranta, J., and Nurmi, J. E. (2013). Psychological control in daily parent-child interactions increases children's negative emotions. *J. Fam. Psychol.* 27, 453–462. doi: 10.1037/a0032891
- Bailey, D. H., Siegler, R. S., and Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Dev. Sci.* 17, 775–785. doi: 10.1111/desc.12155
- Berch, D. B. (2005). Making sense of number sense: implications for children with mathematical disabilities. *J. Learn. Disabil.* 38, 333–339. doi: 10.1177/00222194050380040901
- Bonny, J. W., and Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: Evidence from the preschool years. *J. Exp. Child Psychol.* 114, 375–388. doi: 10.1016/j.jecp.2012.09.015
- Bugden, S., and Ansari, D. (2015). Probing the nature of deficits in the ‘Approximate number System’ in children with persistent developmental dyscalculia. *Dev. Sci.* 19, 817–833. doi: 10.1111/desc.12324
- Bull, R., Espy, K. A., and Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: longitudinal predictors of mathematical achievement at age 7 years. *Dev. Neuropsychol.* 33, 205–228. doi: 10.1080/87565640801982312

- Butterworth, B. (2005a). "Developmental dyscalculia," in *Handbook of Mathematical Cognition*, ed. J. I. D. Campbell (Hove: Psychology Press), 455–467.
- Butterworth, B. (2005b). The development of arithmetical abilities. *J. Child Psychol. Psychiatry* 46, 3–18. doi: 10.1111/j.1469-7610.2004.00374.x
- Campbell, J. I., and Clark, J. M. (1988). An encoding-complex view of cognitive number processing: comment on McCloskey, sokol, and goodman. *J. Exp. Psychol. Gen.* 117, 204–214. doi: 10.1037/0096-3445.117.2.204
- Cantlon, J. F., Davis, S. W., Libertus, M. E., Kahane, J., Brannon, E. M., and Pelphey, K. A. (2011). Inter-parietal white matter development predicts numerical performance in young children. *Learn. Individ. Differ.* 21, 672–680. doi: 10.1016/j.lindif.2011.09.003
- Castro-Canizares, D., Estevez-Perez, N., and Reigosa-Crespo, V. (2009). Contemporary cognitive theories about developmental dyscalculia. *Rev. Neurol.* 49, 143–148.
- Caviola, S., Mammarella, I. C., Lucangeli, D., and Cornoldi, C. (2014). Working memory and domain-specific precursors predicting success in learning written subtraction problems. *Learn. Individ. Differ.* 36, 92–100. doi: 10.1016/j.lindif.2014.10.010
- Chang, T. T., Rosenberg-Lee, M., Metcalfe, A. W., Chen, T., and Menon, V. (2015). Development of common neural representations for distinct numerical problems. *Neuropsychologia* 75, 481–495. doi: 10.1016/j.neuropsychologia.2015.07.005
- Chen, Q., and Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychol.* 148, 163–172. doi: 10.1016/j.actpsy.2014.01.016
- Chu, F. W., and Geary, D. C. (2015). Early numerical foundations of young children's mathematical development. *J. Exp. Child Psychol.* 132, 205–212. doi: 10.1016/j.jecp.2015.01.006
- Cirino, P. T. (2011). The interrelationships of mathematical precursors in kindergarten. *J. Exp. Child Psychol.* 108, 713–733. doi: 10.1016/j.jecp.2010.11.004
- Clark, C. A., Sheffield, T. D., Wiebe, S. A., and Espy, K. A. (2013). Longitudinal associations between executive control and developing mathematical competence in preschool boys and girls. *Child Dev.* 84, 662–677. doi: 10.1111/j.1467-8624.2012.01854.x
- Clearman, J., Klinger, V., and Szűcs, D. (2017). Visuospatial and verbal memory in mental arithmetic. *Q. J. Exp. Psychol.* 70, 1837–1855. doi: 10.1080/17470218.2016.1209534
- Cohen Kadosh, R., Lammertyn, J., and Izard, V. (2008). Are numbers special? an overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Prog. Neurobiol.* 84, 132–147. doi: 10.1016/j.pneurobio.2007.11.001
- Cohen Kadosh, R., and Walsh, V. (2009). Numerical representation in the parietal lobes: Abstract or not abstract? *Behav. Brain Sci.* 32, 313–328. doi: 10.1017/S0140525X09990938
- Crollen, V., Seron, X., and Noël, M.-P. (2011). Is finger-counting necessary for the development of arithmetic abilities? *Front. Psychol.* 2:242. doi: 10.3389/fpsyg.2011.00242
- De Smedt, B., and Boets, B. (2010). Phonological processing and arithmetic fact retrieval: evidence from developmental dyslexia. *Neuropsychologia* 48, 3973–3981. doi: 10.1016/j.neuropsychologia.2010.10.018
- De Visscher, A., and Noël, M. P. (2016). Similarity interference in learning and retrieving arithmetic facts. *Prog. Brain Res.* 227, 131–158. doi: 10.1016/bs.pbr.2016.04.008
- Dehaene, S., and Cohen, L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex* 33, 219–250. doi: 10.1016/S0010-9452(08)70002-9
- Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239
- Desoete, A., Ceulemans, A., De Weerd, F., and Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? findings from a longitudinal study. *Br. J. Educ. Psychol.* 82(Pt 1), 64–81. doi: 10.1348/2044-8279.002002
- Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Dev. Sci.* 11, 650–654. doi: 10.1111/j.1467-7687.2008.00713.x
- Dumontheil, I., and Klingberg, T. (2012). Brain activity during a visuospatial working memory task predicts arithmetical performance 2 years later. *Cereb. Cortex* 22, 1078–1085. doi: 10.1093/cercor/bhr175
- Fazio, L. K., Bailey, D. H., Thompson, C. A., and Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *J. Exp. Child Psychol.* 123, 53–72. doi: 10.1016/j.jecp.2014.01.013
- Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307–314. doi: 10.1016/j.tics.2004.05.002
- Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *J. Educ. Psychol.* 98, 29–43. doi: 10.1037/0022-0663.98.1.29
- Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., et al. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Dev. Psychol.* 46, 1731–1746. doi: 10.1037/a0020662
- Geary, D. C. (2004). Mathematics and learning disabilities. *J. Learn. Disabil.* 37, 4–15. doi: 10.1177/00222194040370010201
- Geary, D. C. (2007). An evolutionary perspective on learning disability in mathematics. *Dev. Neuropsychol.* 32, 471–519. doi: 10.1080/87565640701360924
- Geary, D. C. (2010). "Missouri longitudinal study of mathematical development and disability," in *BJEP Monograph Series II, Number 7-Understanding Number Development and Difficulties*, (London: British Psychological Society), 31–49. doi: 10.1348/97818543370009X1258369932410
- Geary, D. C., Hamson, C. O., and Hoard, M. K. (2000). Numerical and arithmetical cognition: a longitudinal study of process and concept deficits in children with learning disability. *J. Exp. Child Psychol.* 77, 236–263. doi: 10.1006/jecp.2000.2561
- Geary, D. C., Hoard, M. K., and Bailey, D. H. (2012). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *J. Learn. Disabil.* 45, 291–307. doi: 10.1177/0022219410392046
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., and Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Dev.* 78, 1343–1359. doi: 10.1111/j.1467-8624.2007.01069.x
- Geary, D. C., Hoard, M. K., Nugent, L., and Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Dev. Neuropsychol.* 33, 277–299. doi: 10.1080/87565640801982361
- Geary, D. C., and Moore, A. M. (2016). Cognitive and brain systems underlying early mathematical development. *Prog. Brain Res.* 227, 75–103. doi: 10.1016/bs.pbr.2016.03.008
- Geary, D. C., Nicholas, A., Li, Y., and Sun, J. (2017). Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: an eight-year longitudinal study. *J. Educ. Psychol.* 109, 680–693. doi: 10.1037/edu0000159
- Gersten, R., and Chard, D. (1999). Number sense rethinking arithmetic instruction for students with mathematical disabilities. *J. Spec. Educ.* 33, 18–28. doi: 10.1177/002246699903300102
- Gray, S. A., and Reeve, R. A. (2014). Preschoolers' dot enumeration abilities are markers of their arithmetic competence. *PLoS One* 9:e94428. doi: 10.1371/journal.pone.0094428
- Hanich, L. B., Jordan, N. C., Kaplan, D., and Dick, J. (2001). Performance across different areas of mathematical cognition in children with learning difficulties. *J. Educ. Psychol.* 93, 615–626. doi: 10.1037/0022-0663.93.3.615
- Heikkilä, R., Torppa, M., Aro, M., Närhi, V., and Ahonen, T. (2016). Double-deficit hypothesis in a clinical sample: Extension beyond reading. *J. Learn. Disabil.* 49, 546–560. doi: 10.1177/0022219415572895
- Henik, A., Leibovich, T., Naparstek, S., Diesendruck, L., and Rubinsten, O. (2012). Quantities, amounts, and the numerical core system. *Front. Hum. Neurosci.* 5:186. doi: 10.3389/fnhum.2011.00186
- Holloway, I. D., and Ansari, D. (2008). Domain-specific and domain-general changes in children's development of number comparison. *Dev. Sci.* 11, 644–649. doi: 10.1111/j.1467-7687.2008.00712.x
- Hornung, C., Schiltz, C., Brunner, M., and Martin, R. (2014). Predicting first-grade mathematics achievement: The contributions of domain-general cognitive

- abilities, nonverbal number sense, and early number competence. *Front. Psychol.* 5:272. doi: 10.3389/fpsyg.2014.00272
- Hyde, D. C. (2011). Two systems of non-symbolic numerical cognition. *Front. Hum. Neurosci.* 5:150. doi: 10.3389/fnhum.2011.00150
- Hyde, D. C., Bertelelli, I., and Mou, Y. (2016). Approximate numerical abilities and mathematics: insight from correlational and experimental training studies. *Prog. Brain Res.* 227, 335–351. doi: 10.1016/bs.pbr.2016.04.011
- Inglis, M., Attridge, N., Batchelor, S., and Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: but only in children. *Psychon. Bull. Rev.* 18, 1222–1229. doi: 10.3758/s13423-011-0154-1
- Jordan, J. A., Wylie, J., and Mulhern, G. (2015). Mathematics and reading difficulty subtypes: minor phonological influences on mathematics for 5–7-years-old. *Front. Psychol.* 6:221. doi: 10.3389/fpsyg.2015.00221
- Jordan, N. C., Hanich, L. B., and Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Dev.* 74, 834–850. doi: 10.1111/1467-8624.00571
- Jordan, N. C., Kaplan, D., and Hanich, L. B. (2002). Achievement growth in children with learning difficulties in mathematics: findings of a two-year longitudinal study. *J. Educ. Psychol.* 94, 586–597. doi: 10.1037//0022-0663.94.3.586
- Jordan, N. C., Kaplan, D., Locuniak, M. N., and Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learn. Disabil. Res. Pract.* 22, 36–46. doi: 10.1111/j.1540-5826.2007.00229.x
- Jordan, N. C., Kaplan, D., Ramineni, C., and Locuniak, M. N. (2009). Early math matters: kindergarten number competence and later mathematics outcomes. *Dev. Psychol.* 45, 850–867. doi: 10.1037/a0014939
- Kamawar, D., LeFevre, J., Bisanz, J., Fast, L., Skwarchuk, S., Smith-Chant, B., et al. (2010). Knowledge of counting principles: How relevant is order irrelevance? *J. Exp. Child Psychol.* 105, 138–145. doi: 10.1016/j.jecp.2009.08.004
- Karmiloff-Smith, A. (2015). An alternative to domain-general or domain-specific frameworks for theorizing about human evolution and ontogenesis. *AIMS Neurosci.* 2, 91–104. doi: 10.3934/Neuroscience.2015.2.91
- Kaufmann, L., Mazzocco, M. M., Dowker, A., von Aster, M., Gobel, S. M., Grabner, R. H., et al. (2013). Dyscalculia from a developmental and differential perspective. *Front. Psychol.* 4:516. doi: 10.3389/fpsyg.2013.00516
- Klein, E., Moeller, K., and Willmes, K. (2013). A neural disconnection hypothesis on impaired numerical processing. *Front. Hum. Neurosci.* 7:663. doi: 10.3389/fnhum.2013.00663
- Knievel, J., Petermann, F., and Daseking, M. (2011). Welche Vorläuferdefizite weisen Kinder mit einer kombinierten Rechtschreib- und Rechenschwäche auf? *Diagnostica* 57, 212–224. doi: 10.1026/0012-1924/a000054
- Kolkman, M. E., Kroesbergen, E. H., and Leseman, P. P. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learn. Instr.* 25, 95–103. doi: 10.1016/j.learninstruc.2012.12.001
- Krajewski, K., and Schneider, W. (2009). Early development of quantity to number-word linkage as a precursor of mathematical school achievement and mathematical difficulties: findings from a four-year longitudinal study. *Learn. Instr.* 19, 513–526. doi: 10.1016/j.learninstruc.2008.10.002
- Kucian, K., and Kaufmann, L. (2009). A developmental model of number representation. *Behav. Brain Sci.* 32, 340–341. doi: 10.1017/S0140525X09990069
- Kucian, K., and von Aster, M. (2015). Developmental dyscalculia. *Eur. J. Pediatr.* 174, 1–13. doi: 10.1007/s00431-014-2455-7
- Kuhn, J., Ise, E., Raddatz, J., Schwenk, C., and Dobel, C. (2016). Basic numerical processing, calculation, and working memory in children with dyscalculia and/or ADHD symptoms. *Z. Kinder Jugendpsychiatr. Psychother.* 44, 365–375. doi: 10.1024/1422-4917/a000450
- Kuhn, J.-T., and Holling, H. (2014). Number sense or working memory? The effect of two computer-based trainings on mathematical skills in elementary school. *Adv. Cogn. Psychol.* 10, 59–67. doi: 10.5709/acp-0157-2
- Landerl, K., and Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. *J. Exp. Child Psychol.* 103, 546–565. doi: 10.1016/j.jecp.2008.12.006
- Lauer, J. E., and Lourenco, S. F. (2016). Spatial processing in infancy predicts both spatial and mathematical aptitude in childhood. *Psychol. Sci.* 27, 1291–1298. doi: 10.1177/0956797616655977
- LeFevre, J., Berrigan, L., Vendetti, C., Kamawar, D., Bisanz, J., Skwarchuk, S., et al. (2013). The role of executive attention in the acquisition of mathematical skills for children in grades 2 through 4. *J. Exp. Child Psychol.* 114, 243–261. doi: 10.1016/j.jecp.2012.10.005
- LeFevre, J. A. (2016). Numerical cognition: Adding it up. *Can. J. Exp. Psychol.* 70, 3–11. doi: 10.1037/cep0000062
- LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Dev.* 81, 1753–1767. doi: 10.1111/j.1467-8624.2010.01508.x
- Lembke, E., and Foegen, A. (2009). Identifying early numeracy indicators for kindergarten and first-grade students. *Learn. Disabil. Res. Pract.* 24, 12–20. doi: 10.1111/j.1540-5826.2008.01273.x
- Libertus, M. E., Feigenson, L., and Halberda, J. (2013a). Is approximate number precision a stable predictor of math ability? *Learn. Individ. Differ.* 25, 126–133. doi: 10.1016/j.lindif.2013.02.001
- Libertus, M. E., Feigenson, L., and Halberda, J. (2013b). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *J. Exp. Child Psychol.* 116, 829–838. doi: 10.1016/j.jecp.2013.08.003
- Lindskog, M., Winman, A., and Juslin, P. (2014). The association between higher education and approximate number system acuity. *Front. Psychol.* 5:462. doi: 10.3389/fpsyg.2014.00462
- Locuniak, M. N., and Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *J. Learn. Disabil.* 41, 451–459. doi: 10.1177/0022219408321126
- Lyons, I. M., Ansari, D., and Beilock, S. L. (2012). Symbolic estrangement: evidence against a strong association between numerical symbols and the quantities they represent. *J. Exp. Psychol. Gen.* 141, 635–641. doi: 10.1037/a0027248
- Mammarella, I., Caviola, S., Cornoldi, C., and Lucangeli, D. (2013). Mental additions and verbal-domain interference in children with developmental dyscalculia. *Res. Dev. Disabil.* 34, 2845–2855. doi: 10.1016/j.ridd.2013.05.044
- Mazzocco, M. M., Feigenson, L., and Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Dev.* 82, 1224–1237. doi: 10.1111/j.1467-8624.2011.01608.x
- Mazzocco, M. M., and Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Ann. Dyslexia* 53, 218–253. doi: 10.1007/s11881-003-0011-7
- Mazzocco, M. M., and Thompson, R. E. (2005). Kindergarten predictors of math learning disability. *Learn. Disabil. Res. Pract.* 20, 142–155. doi: 10.1111/j.1540-5826.2005.00129.x
- Methe, S. A., Hintze, J. M., and Floyd, R. G. (2008). Validation and decision accuracy of early numeracy skill indicators. *Sch. Psychol. Rev.* 37, 359–373.
- Moeller, K., Willmes, K., and Klein, E. (2015). A review on functional and structural brain connectivity in numerical cognition. *Front. Hum. Neurosci.* 9:227. doi: 10.3389/fnhum.2015.00227
- Moore, A. M., vanMarle, K., and Geary, D. C. (2016). Kindergartners' fluent processing of symbolic numerical magnitude is predicted by their cardinal knowledge and implicit understanding of arithmetic 2 years earlier. *J. Exp. Child Psychol.* 150, 31–47. doi: 10.1016/j.jecp.2016.05.003
- Morgan, P. L., Farkas, G., Hillemeier, M. M., and Maczuga, S. (2016). Who is at risk for persistent mathematics difficulties in the United States? *J. Learn. Disabil.* 49, 305–319. doi: 10.1177/0022219414553849
- Mórocz, I. Á., Janos, F., van Gelderen, P., Manor, D., Karni, A., Breznitz, Z., et al. (2012). Time-resolved and spatio-temporal analysis of complex cognitive processes and their role in disorders like developmental dyscalculia. *Int. J. Imaging Syst. Technol.* 22, 81–96. doi: 10.1002/ima.22009
- Murphy, M. M., Mazzocco, M. M., Hanich, L. B., and Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *J. Learn. Disabil.* 40, 458–478. doi: 10.1177/00222194070400050901
- Namkung, J. M., and Fuchs, L. S. (2016). Cognitive predictors of calculations and number line estimation with whole numbers and fractions among at-risk students. *J. Educ. Psychol.* 108, 214–228. doi: 10.1037/edu0000055
- National Mathematics Advisory Panel (2008). *Foundations for Success: Final Report of the National Mathematics Advisory Panel*. Washington, DC: United States Department of Education.
- Noël, M., and Rousselle, L. (2011). Developmental changes in the profiles of dyscalculia: an explanation based on a double exact-and-approximate number

- representation model. *Front. Hum. Neurosci.* 5:165. doi: 10.3389/fnhum.2011.00165
- Nys, J., Ventura, P., Fernandes, T., Querido, L., Leybaert, J., and Content, A. (2013). Does math education modify the approximate number system? A comparison of schooled and unschooled adults. *Trends Neurosci. Educ.* 2, 13–22. doi: 10.1016/j.tine.2013.01.001
- Obersteiner, A., Reiss, K., and Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learn. Instr.* 23, 125–135. doi: 10.1016/j.learninstruc.2012.08.004
- Park, J., and Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychol. Sci.* 24, 2013–2019. doi: 10.1177/0956797613482944
- Park, J., and Brannon, E. M. (2014). Improving arithmetic performance with number sense training: an investigation of underlying mechanism. *Cognition* 133, 188–200. doi: 10.1016/j.cognition.2014.06.011
- Passolunghi, M. C., Mammarella, I. C., and Altoe, G. (2008). Cognitive abilities as precursors of the early acquisition of mathematical skills during first through second grades. *Dev. Neuropsychol.* 33, 229–250. doi: 10.1080/87565640801982320
- Peng, P., Namkung, J., Barnes, M., and Sun, C. (2016a). A meta-analysis of mathematics and working memory: Moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *J. Educ. Psychol.* 108, 455–473. doi: 10.1037/edu0000079
- Peng, P., Namkung, J. M., Fuchs, D., Fuchs, L. S., Patton, S., Yen, L., et al. (2016b). A longitudinal study on predictors of early calculation development among young children at risk for learning difficulties. *J. Exp. Child Psychol.* 152, 221–241. doi: 10.1016/j.jecp.2016.07.017
- Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends Cogn. Sci.* 14, 542–551. doi: 10.1016/j.tics.2010.09.008
- Piazza, M., Pica, P., Izard, V., Spelke, E. S., and Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychol. Sci.* 24, 1037–1043. doi: 10.1177/0956797612464057
- Powell, S. R., and Fuchs, L. S. (2012). Early numerical competencies and students with mathematics difficulty. *Focus Except Child* 44, 1–16. doi: 10.17161/fec.v44i5.6686
- Praet, M., Titeca, D., Ceulemans, A., and Desoete, A. (2013). Language in the prediction of arithmetics in kindergarten and grade 1. *Learn. Individ. Differ.* 27, 90–96. doi: 10.1016/j.lindif.2013.07.003
- Raghubar, K. P., and Barnes, M. A. (2017). Early numeracy skills in preschool-aged children: a review of neurocognitive findings and implications for assessment and intervention. *Clin. Neuropsychol.* 31, 328–350. doi: 10.1080/13854046.2016.1259387
- Rapin, I. (2016). Dyscalculia and the calculating brain. *Pediatr. Neurol.* 61, 11–20. doi: 10.1016/j.pediatrneurol.2016.02.007
- Reeve, R. A., Paul, J. M., and Butterworth, B. (2015). Longitudinal changes in young children's 0-100 to 0-1000 number-line error signatures. *Front. Psychol.* 6:647. doi: 10.3389/fpsyg.2015.00647
- Rousselle, L., and Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: a comparison of symbolic vs non-symbolic number magnitude processing. *Cognition* 102, 361–395. doi: 10.1016/j.cognition.2006.01.005
- Sasanguie, D., de Smedt, B., and Reynvoet, B. (2015). Evidence for distinct magnitude systems for symbolic and non-symbolic number. *Psychol. Res.* 81, 231–242. doi: 10.1007/s00426-015-0734-1
- Sasanguie, D., Gobel, S. M., Moll, K., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *J. Exp. Child Psychol.* 114, 418–431. doi: 10.1016/j.jecp.2012.10.012
- Sasanguie, D., Van den Bussche, E., and Reynvoet, B. (2012). Predictors for mathematics achievement? Evidence from a longitudinal study. *Mind Brain Educ.* 6, 119–128. doi: 10.1111/j.1751-228X.2012.01147.x
- Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science* 20, e12372. doi: 10.1111/desc.12372
- Schulte-Körne, G. (2014). Spezifische Lernstörungen. *Z. Kinder Jugendpsychiatr. Psychother.* 42, 369–374. doi: 10.1024/1422-4917/a000312
- Schwenk, C., Sasanguie, D., Kuhn, J.-T., Kempe, S., Doeblner, P., and Holling, H. (2017). (Non)symbolic magnitude processing in mathematically low-achieving and dyscalculic children: a meta-analysis. *Res. Dev. Disabil.* 64, 152–167. doi: 10.1016/j.ridd.2017.03.003
- Siegler, R. S., Duncan, G. J., Davis-Kean, P. E., Duckworth, K., Claessens, A., Engel, M., et al. (2012). Early predictors of high school mathematics achievement. *Psychol. Sci.* 23, 691–697. doi: 10.1177/0956797612440101
- Simmons, F. R., and Singleton, C. (2008). Do weak phonological representations impact on arithmetic development? A review of research into arithmetic and dyslexia. *Dyslexia* 14, 77–94. doi: 10.1002/dys.341
- Skagerlund, K., and Traff, U. (2016). Number processing and heterogeneity of developmental dyscalculia: subtypes with different cognitive profiles and deficits. *J. Learn. Disabil.* 49, 36–50. doi: 10.1177/0022219414522707
- Sowinski, C., LeFevre, J. A., Skwarchuk, S. L., Kamawar, D., Bisanz, J., and Smith-Chant, B. (2015). Refining the quantitative pathway of the pathways to mathematics model. *J. Exp. Child Psychol.* 131, 73–93. doi: 10.1016/j.jecp.2014.11.004
- Swanson, H. L. (2011). Working memory, attention, and mathematical problem solving: a longitudinal study of elementary school children. *J. Educ. Psychol.* 103, 821–837. doi: 10.1037/a0025114
- Swanson, H. L. (2012). Cognitive profile of adolescents with math disabilities: Are the profiles different from those with reading disabilities? *Child Neuropsychol.* 18, 125–143. doi: 10.1080/09297049.2011.589377
- Szkudlarek, E., and Brannon, E. M. (2017). Does the approximate number system serve as a foundation for symbolic mathematics? *Lang. Learn. Dev.* 13, 171–190. doi: 10.1080/15475441.2016.1263573
- Szűcs, D. (2016). Subtypes and comorbidity in mathematical learning disabilities: multidimensional study of verbal and visual memory processes is key to understanding. *Prog. Brain Res.* 227, 277–304. doi: 10.1016/bs.pbr.2016.04.027
- Szűcs, D., Devine, A., Soltesz, F., Nobes, A., and Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex* 49, 2674–2688. doi: 10.1016/j.cortex.2013.06.007
- Szűcs, D., Devine, A., Soltesz, F., Nobes, A., and Gabriel, F. (2014). Cognitive components of a mathematical processing network in 9-year-old children. *Dev. Sci.* 17, 506–524. doi: 10.1111/desc.12144
- Szűcs, D., and Myers, T. (2017). A critical analysis of design, facts, bias and inference in the approximate number system training literature: a systematic review. *Trends Neurosci. Educ.* 6, 187–203. doi: 10.1016/j.tine.2016.11.002
- Toll, S. W., Kroesbergen, E. H., and Van Luit, J. E. (2016). Visual working memory and number sense: testing the double deficit hypothesis in mathematics. *Br. J. Educ. Psychol.* 86, 429–445. doi: 10.1111/bjep.12116
- Traeff, U., Olsson, L., Östergren, R., and Skagerlund, K. (2017). Heterogeneity of developmental dyscalculia: cases with different deficit profiles. *Front. Psychol.* 7:2000. doi: 10.3389/fpsyg.2016.02000
- Vandenbergh, R., Molenberghs, P., and Gillebert, C. R. (2012). Spatial attention deficits in humans: the critical role of superior compared to inferior parietal lesions. *Neuropsychologia* 50, 1092–1103. doi: 10.1016/j.neuropsychologia.2011.12.016
- VanDerHeyden, A. M., Broussard, C., and Cooley, A. (2006). Further development of measures of early math performance for preschoolers. *J. Sch. Psychol.* 44, 533–553. doi: 10.1016/j.jsp.2006.07.003
- Vandervort, L. (2017). The origin of mathematics and number sense in the cerebellum: with implications for finger counting and dyscalculia. *Cerebellum Ataxias* 4:12. doi: 10.1186/s40673-017-0070-x
- Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., and Newcombe, N. S. (2017). Vi. discussion and implications: How early spatial skills predict later spatial and mathematical skills. *Monogr. Soc. Res. Child Dev.* 82, 89–109. doi: 10.1111/mono.12285
- von Aster, M. G., and Shalev, R. S. (2007). Number development and developmental dyscalculia. *Dev. Med. Child Neurol.* 49, 868–873. doi: 10.1111/j.1469-8749.2007.00868.x
- Willcutt, E. G., Petrill, S. A., Wu, S., Boada, R., Defries, J. C., Olson, R. K., et al. (2013). Comorbidity between reading disability and math disability: Concurrent psychopathology, functional impairment, and neuropsychological functioning. *J. Learn. Disabil.* 46, 500–516. doi: 10.1177/0022219413477476
- Wilson, A. J., and Dehaene, S. (2007). "Number sense and developmental dyscalculia," in *Human Behavior, Learning, and the Developing Brain: Atypical*

- Development*, eds D. Coch, G. Dawson, and K. W. Fischer (New York, NY: Guilford Press), 212–238.
- Wolf, M., and Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *J. Educ. Psychol.* 91, 415–438. doi: 10.1037/0022-0663.91.3.415
- Wong, T. T., Ho, C. S., and Tang, J. (2017). Defective number sense or impaired access? Differential impairments in different subgroups of children with mathematics difficulties. *J. Learn. Disabil.* 50, 49–61. doi: 10.1177/0022219415588851
- World Health Organization [WHO] (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. Geneva: World Health Organization.
- Yang, J., Han, H., Chui, D., Shen, Y., and Wu, J. (2012). Prominent activation of the intraparietal and somatosensory areas during angle discrimination by intra-active touch. *Hum. Brain Mapp.* 33, 2957–2970. doi: 10.1002/hbm.21419
- Zamarian, L., Ischebeck, A., and Delazer, M. (2009). Neuroscience of learning arithmetic—evidence from brain imaging studies. *Neurosci. Biobehav. Rev.* 33, 909–925. doi: 10.1016/j.neubiorev.2009.03.005
- Zebian, S., and Ansari, D. (2012). Differences between literates and illiterates on symbolic but not nonsymbolic numerical magnitude processing. *Psychon. Bull. Rev.* 19, 93–100. doi: 10.3758/s13423-011-0175-9
- Zhang, X., Koponen, T., Rasanen, P., Aunola, K., Lerkkanen, M. K., and Nurmi, J. E. (2014). Linguistic and spatial skills predict early arithmetic development via counting sequence knowledge. *Child Dev.* 85, 1091–1107. doi: 10.1111/cdev.12173

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

Copyright © 2018 Siemann and Petermann. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Relation Between Mathematical Performance, Math Anxiety, and Affective Priming in Children With and Without Developmental Dyscalculia

Karin Kucian^{1,2,3*}, Isabelle Zuber¹, Juliane Kohn⁴, Nadine Poltz⁴, Anne Wyschkon⁴, Günter Esser⁵ and Michael von Aster^{1,2,3,6}

¹ Center for MR-Research, University Children's Hospital, Zurich, Switzerland, ² Children's Research Center, University Children's Hospital, Zurich, Switzerland, ³ Neuroscience Center Zurich, University of Zurich, ETH Zurich, Zurich, Switzerland, ⁴ Department of Psychology, University of Potsdam, Potsdam, Germany, ⁵ Academy for Psychotherapy and Intervention Research, University of Potsdam, Potsdam, Germany, ⁶ Clinic for Child and Adolescent Psychiatry, German Red Cross Hospitals, Berlin, Germany

OPEN ACCESS

Edited by:

Bert De Smedt,
KU Leuven, Belgium

Reviewed by:

Brenda R. J. Jansen,
University of Amsterdam, Netherlands
Kinga Morsanyi,
Queen's University Belfast,
United Kingdom

*Correspondence:

Karin Kucian
karin.kucian@kispi.uzh.ch

Specialty section:

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

Received: 10 October 2017

Accepted: 16 February 2018

Published: 26 April 2018

Citation:

Kucian K, Zuber I, Kohn J, Poltz N,
Wyschkon A, Esser G and
von Aster M (2018) Relation Between
Mathematical Performance, Math
Anxiety, and Affective Priming in
Children With and Without
Developmental Dyscalculia.
Front. Psychol. 9:263.
doi: 10.3389/fpsyg.2018.00263

Many children show negative emotions related to mathematics and some even develop mathematics anxiety. The present study focused on the relation between negative emotions and arithmetical performance in children with and without developmental dyscalculia (DD) using an affective priming task. Previous findings suggested that arithmetic performance is influenced if an affective prime precedes the presentation of an arithmetic problem. In children with DD specifically, responses to arithmetic operations are supposed to be facilitated by both negative and mathematics-related primes (*=negative math priming effect*). We investigated mathematical performance, math anxiety, and the domain-general abilities of 172 primary school children (76 with DD and 96 controls). All participants also underwent an affective priming task which consisted of the decision whether a simple arithmetic operation (addition or subtraction) that was preceded by a prime (positive/negative/neutral or mathematics-related) was true or false. Our findings did not reveal a *negative math priming effect* in children with DD. Furthermore, when considering accuracy levels, gender, or math anxiety, the *negative math priming effect* could not be replicated. However, children with DD showed more math anxiety when explicitly assessed by a specific math anxiety interview and showed lower mathematical performance compared to controls. Moreover, math anxiety was equally present in boys and girls, even in the earliest stages of schooling, and interfered negatively with performance. In conclusion, mathematics is often associated with negative emotions that can be manifested in specific math anxiety, particularly in children with DD. Importantly, present findings suggest that in the assessed age group, it is more reliable to judge math anxiety and investigate its effects on mathematical performance explicitly by adequate questionnaires than by an affective math priming task.

Keywords: developmental dyscalculia, mathematics, affective priming, calculation, arithmetic, anxiety, gender, children

INTRODUCTION

Mathematical skills are vital for everyday life and deficits in mathematical performance have negative effects in many domains like education, profession and daily routine. It is commonly known that many children have negative attitudes and emotions toward mathematics (Dowker et al., 2016). In some children the negative emotions toward mathematics may evoke severe anxiety, and as a consequence, these children often avoid mathematical activities.

The literature has proposed different definitions of mathematics anxiety, however, common to most is the observation that dealing with mathematics may evoke a negative emotional response in some people (Suárez-Pellicioni et al., 2016). Mathematics anxiety involves feelings of tension and interferes with mathematical performance (Dowker et al., 2016). Mathematics anxiety is certainly a significant problem that appears to increase with age during childhood. According to the Organization for Economic Co-operation and Development (OECD), 31% of 15-year-old students reported feeling nervous when solving a math problem and as many as 59% indicated that they were worried about math classes (OECD, 2013). Apart from aspects like gender, age and culture affecting mathematics anxiety, research has shown that emotional factors, such as general anxiety or self-esteem play an important role too (Orly Rubinsten and Tannock, 2010; Dowker et al., 2016). Environmental and genetic factors have also been discussed. As associations between mathematics anxiety and achievement have been validated, it is assumed that children with learning disabilities in mathematics show higher levels of mathematics anxiety (Wu et al., 2014). Hence, children suffering from math learning disorders, such as developmental dyscalculia (DD), are of particular interest when investigating these relations. With a prevalence rate of between 3 and 6%, children suffering from DD are clearly not a rare exception (Shalev et al., 2000). DD is a heterogeneous learning impairment affecting numerical and/or arithmetic functioning on the behavioral, psychological and neuronal levels (reviewed by Kucian and von Aster, 2015). Hereditary and environmental factors are presumed to represent possible causes, and children affected from DD report problems with counting, magnitude processing, arithmetic but also more general competences such as working memory or attentional processes. Furthermore, children with DD often suffer from additional psychiatric disorders like depression or anxieties. Anxiety is especially present in the context of mathematics and is associated with stress (reviewed by Dowker et al., 2016).

Despite their relatively high occurrence and significant importance, comparatively little research has been conducted on the interaction between low performance in mathematics and negative emotions. The topic has received increasing attention, yet much remains unexplained and contradictory findings have been reported. Of particular interest is the relation between cognitive abilities and emotional factors and attitudes in mathematical performance (e.g., Dowker et al., 2012). However, the direction of causation is undefined. On the one hand, it is possible that having high mathematics anxiety leads to greater avoidance tendencies in situations that involve mathematics,

resulting in less practice and hence lower achievement. On the other hand, it is also plausible that poor mathematical performance promotes mathematics anxiety. Moreover, working memory capacity has been shown to be lower in highly math-anxious subjects (e.g., Mammarella et al., 2015). Interestingly, the authors reported that children with math anxiety are specifically impaired in verbal working memory, whereas children with DD showed specific deficits in visuospatial working memory. Hence, children with math anxiety or with DD may fail in math due to different underlying cognitive impairments in working memory. Although it is unclear to what extent mathematics anxiety causes mathematical difficulties or vice versa, there is conclusive proof that math anxiety interferes with mathematical performance, especially with tasks requiring working memory. The most prominent theory explains this relationship by worrying intrusive thoughts involved in math anxiety that consume attentional working memory resources, such that fewer resources are available for numerical cognition (reviewed by Suárez-Pellicioni et al., 2016). Apart from the general lack of research, particularly little is known about the relationship between mathematics anxiety and performance in young children, as previous studies mostly included older children, adolescents or adults. The understanding of early development, however, is crucial in order to prevent mathematics anxiety and negative emotions in the context of mathematical performance (Wu et al., 2012).

One possible strategy for studying the link between mathematical performance and emotions is by the use of priming tasks. Priming tasks are implicit measures and assess evaluations that are activated automatically after the presentation of a stimulus (Krause et al., 2012). Hence, in priming paradigms, resulting effects are mainly caused by response activation processes (De Houwer et al., 2009). The type of priming which is relevant in the context of mathematics anxiety is affective priming, which, in accordance with standard priming paradigms, consists of a stimulus (prime) and a response to a target. Importantly, in affective priming tasks, the affective relation between prime and target is manipulated since the valence of the prime stimulus is either positive, negative or neutral (Hermans et al., 1994). The idea in affective priming tasks is that “participants are faster at evaluating a target stimulus if a previously presented prime stimulus has the same valence compared to a condition in which a prime stimulus of the opposite valence is shown” (Werner and Rothermund, 2013, p. 119). Hence, if the prime-target pair is of the same valence (e.g., positive prime–positive target), processing is facilitated and results in shorter reaction times, whereas if it is of different valence (e.g., positive prime–negative target), processing is inhibited and followed by longer reaction times (Hermans et al., 1994).

Accordingly, to further elucidate the relationship between arithmetic, emotions and low achievement, Rubinsten and Tannock (2010) investigated the effects of mathematics anxiety on numerical processing using a novel affective priming task. Their task differed from that typically utilized in standard affective priming procedures, in that not only positive, negative and neutral primes but also mathematics-related ones were included. The assumption was that this arithmetic-affective

priming task acts as an indirect measure of mathematics anxiety. The sample consisted of 23 participants (12 children with DD and 11 control children) that were all above grade 4. Children had to complete a priming task where a priming word was followed by an arithmetic operation. They then had to indicate whether the arithmetic operation was true or false. As mentioned above, in priming tasks the priming word often cannot be ignored by participants, which is why it interferes with arithmetic performance. The priming words presented were either with a positive, negative or neutral affect or with some relation to mathematics. The arithmetic operations included were single-digit additions, subtractions, multiplications or divisions. As hypothesized, a direct link appeared between emotions (primes) and the arithmetic operations, and this association was different in children with DD and controls. Precisely, children with DD responded faster when the preceding priming word was negative or mathematics-related. This is in line with the assumption of affective priming paradigms that if the prime-target pair is of the same valence, processing is facilitated and hence faster. This implies that for DD children, mathematics-related primes are as negatively attributed as negative primes themselves. In the control children a reversed pattern was observed since mathematic related primes inhibited processing. Based on their findings, Rubinsten and Tannock concluded that a negative math priming effect exists in children with DD and that the arithmetic-affective priming task could be used as an indirect measure of mathematics anxiety in these children.

The current study aimed to reinvestigate the finding that in children with DD mathematics-related and negative primes have a similar effect on performance, particularly a facilitative influence on arithmetical processing. This analogous influence of negative affective and mathematics-related primes is henceforth referred to as the *negative math priming effect*. To enable an in-depth study of the relation between mathematics, emotions and performance, a large number of children ranging in age from 7.3 to 11.3 years was examined by detailed neuropsychological assessments and an adapted version of the priming task by Rubinsten and Tannock (2010). Notably, in addition to the indirect measure of mathematics anxiety, we also included a direct measure, namely the Math-Anxiety-Interview (please see section Cognitive assessments). Accordingly, the present study addresses the above-mentioned general lack of research in younger children and provides new insights into direct and indirect measures of the relation between mathematical performance and emotions.

MATERIALS AND METHODS

Subjects

A total of 183 children were recruited and their parents agreed to participate. The aim was that approximately half of the children had a diagnosis of DD and the other half with typical development to achieve equally sized groups. Of the 183 children, 172 children (aged 7.3–11.3 years, mean 8.6 years, 69.8% female) met the general inclusion criteria and hence comprised the final study sample. 76 children (44.2%) further met the criteria for DD, the other 96 children served as control children (CC) (see

also **Table 1** for demographic and behavioral data). The children were recruited in Germany (Berlin, Potsdam) and Switzerland (Zurich).

General Inclusion/Exclusion Criteria

Of the originally 183 children, a total of 11 children (6%) were excluded from the data analyses due to the following criteria: Three children were excluded because their Intelligence Quotient (IQ) was above 120. Seven children were excluded due to psychiatric diagnoses. One child was excluded due to undefined group membership (with mathematical performance $T = 40.44$, see criteria for classification of DD).

Parents gave written consent and children received a voucher for their participation. The study was approved by the local ethics committee based on guidelines from the World Medical Association's Declaration of Helsinki (WMA, 2002).

Classification of DD

Classification of DD was based on the diagnostic criteria of the DSM-V (APA, 2013): criterion (A) severe math problems for more than 6 months; criterion (B) Low achievement scores in one or more standardized mathematical tests (1–2.5 SD below the population mean for age); criterion (C) mathematical difficulties readily apparent in the early school years; criterion (D) not attributable to intellectual disabilities ($IQ > 70$), global developmental delay, hearing or vision disorders, or neurological or motor disorders. In the present study, mathematical performance was assessed in all children by a careful selection of standardized neuropsychological tests particularly designed for the clinical assessment of DD (for detailed description see section Cognitive Assessments). To be classified as having DD, the mean T -value of mathematical performance had to be lower than 40 (< 1 SD). In addition, general intelligence had to be in the normal range ($85 < IQ < 114$, $N = 173$; or marginally above $IQ 115–120$, $N = 7$) with no evidence of any psychiatric disease. According to DSM-V, the discrepancy between mathematical performance and individual IQ is not a requirement for diagnosis, nevertheless, 78% of the DD children showed a discrepancy between mathematical performance and IQ of more than one standard deviation ($N = 59$, mean discrepancy = 14.3 t -value), and 22% showed a lower discrepancy between both measures ($N = 17$, mean discrepancy = 7.1 t -value).

Cognitive Assessments

Intelligence

Estimated intelligence was measured by the mean of different IQ subtests. Mean IQ of children recruited in Switzerland was based on six subtests of the standardized Wechsler Intelligence Scale for Children (WISC-IV) test battery (block design, similarities, digit span, picture concepts, vocabulary, and arithmetic) (Petermann and Petermann, 2007). The mean IQ of children recruited in Germany was based on four subtests, including two of the WISC-IV (block design, similarities) and two of the test battery "Basisdiagnostik Umschriebener Entwicklungsstörungen im Grundschulalter" (BUEGA) (Esser et al., 2008).

TABLE 1 | Demographic and behavioral data of the sample.

	Total	DD	CC	Statistics
Subjects (N)	172	76	96	
Age (years) <i>M</i> (<i>SD</i>)	8.59 (0.95)	9.04 (1.03)	8.23 (0.71)	$U = 1863.5, z = -5.51, p < 0.001$
Gender (male/female)	52/120	22/54	30/66	$\chi^2(1, N = 172) = 0.11, n.s.$
Intelligence (IQ) ^a <i>M</i> (<i>SD</i>)	101.06 (7.69)	96.53 (5.73)	104.65 (7.16)	$U = 1402.5, z = -6.92, p < 0.001$
Mathematical performance (T) ^b <i>M</i> (<i>SD</i>)	43.83 (9.39)	34.99 (3.65)	50.83 (6.00)	$U < 0.001, z = -11.25, p < 0.001$
Math anxiety (intensity) ^c <i>M</i> (<i>SD</i>)	3.24 (2.52)	4.37 (2.41)	2.35 (2.24)	$U = 1909, z = -5.37, p < 0.001$
Arithmetic fluency (T) ^d <i>M</i> (<i>SD</i>)	41.60 (10.13)	33.35 (5.71)	48.13 (7.84)	$t(169) = 14.29, p < 0.001$
Addition (% correct) ^e <i>M</i> (<i>SD</i>)	81.66 (20.72)	68.31 (23.65)	92.30 (8.53)	$U = 1058, z = -7.30, p < 0.001$
Subtraction (% correct) ^e <i>M</i> (<i>SD</i>)	71.53 (23.57)	54.51 (22.76)	85.11 (13.08)	$U = 812.5, z = -8.09, p < 0.001$
Number line (% deviation) ^e <i>M</i> (<i>SD</i>)	4.73 (2.64)	5.77 (3.03)	3.91 (1.92)	$U = 1977, z = -4.06, p < 0.001$
Number line (T) ^f <i>M</i> (<i>SD</i>)	55.95 (14.65)	47.19 (12.71)	62.87 (12.2)	$U = 1226, z = -7.47, p < 0.001$
Working memory (items) ^g <i>M</i> (<i>SD</i>)	4.05 (1.65)	3.77 (1.58)	4.28 (1.68)	$U = 3184.5, z = -1.12, n.s.$

^aMean IQ based on 4 subtests [verbal IQ and matrices test of BUEGA, block design and similarities subtest of WISC-IV ($n = 153$)], or based on 6 subtests of the WISC-IV [block design, similarities, digit span, picture concepts, vocabulary, arithmetic ($n = 19$)].

^bMean mathematical performance based on 4 subtests (addition and subtraction of HRT, Zahlenstrahl II of ZAREKI-R, Rechentest of BUEGA) ($n = 153$), or based on addition and subtraction of HRT and ZAREKI-R ($n = 19$) in *T*-values.

^cMean intensity of math anxiety assessed by the math anxiety interview (MAI); 0 = no math anxiety, 10 = very high math anxiety.

^dMean of addition and subtraction of the HRT in *T*-values.

^eBased on number line task. The percentage of correctly solved addition or subtraction problems are listed. Moreover, the percentage of the deviation between the exact location on the number line and the marked location of the child is indicated.

^fBased on subtests number line I and II of ZAREKI-R in *T*-values.

^gBased on maximum number of correctly recalled items of the Corsi-Suppression test.

Mathematical Performance

The main mathematical and numerical performance for children recruited in Switzerland was assessed using the two subtests of the Heidelberger Rechentest (HRT) (addition, subtraction) (Haffner et al., 2005) and the standardized Neuropsychological Test Battery for Number Processing and Calculation in Children (ZAREKI-R) (von Aster et al., 2006). This neuropsychological battery examines basic skills in calculation and arithmetic and aims to identify and characterize the profile of mathematical abilities in children with DD from the 1st to 4th grade level. It is composed of 11 subtests, such as reverse counting, subtraction, number reading, dictating, visual estimation of quantities, and digit span forward and backward. Mean number processing for children recruited in Germany was assessed using four subtests, namely two of the Heidelberger Rechentest (HRT) (addition, subtraction) (Haffner et al., 2005), the number line task II of the ZAREKI-R and the calculation test of the BUEGA (Esser et al., 2008). The calculation test of the BUEGA evaluates by text problems, which are illustrated in pictures, the knowledge of number comparisons, magnitudes, sizes, and the understanding and use of the four basic arithmetical operations. Criteria for DD for both Swiss and German children were met if a child's performance was below a mean *T*-value of 40.

Mathematics Anxiety

Mathematics anxiety was assessed by the Math-Anxiety-Interview for German speaking primary school children (MAI), which is a valid and reliable measure for the assessment of math anxiety as demonstrated by a Cronbach's alpha of 0.90 (Kohn et al., 2013). The MAI combines two different types of questions while four math related situations are verbally and

pictorially presented (1st on the eve of a math test, 2nd math homework, 3rd math class, and 4th everyday/shopping). The child is initially asked to rate the intensity of his or her anxiety concerning the presented situation by an anxiety thermometer from 0 to 10. In a second step, the different components of anxiety (affective, cognitive, behavioral and physiological) are explored. The child is asked to estimate, to what extent specific statements apply to the particular situation, e.g., "I cannot get a word out." For the present study we have chosen the mean math anxiety intensity associated with all four situations which provides a valid and reliable measure from 0 = no anxiety to 10 = very strong math anxiety in primary school children.

Arithmetic Fluency

Arithmetic fluency was evaluated using the addition and subtraction subtests of the Heidelberger Rechentest (HRT). In this test, a list of 40 addition or subtraction tasks is presented to the child and he/she is asked to solve as many problems as possible within 2 min.

Number Line Performance

The spatial representation of numbers was measured by a paper-and-pencil number line task (Kucian et al., 2011). Children had to indicate with a pencil on a left-to-right oriented number line from 0 to 100 the location of 20 Arabic digits, the results of 20 additions and 20 subtractions, and the estimated number of 10 different dot arrays. The accuracy was measured by calculating the percentage distance from the marked to the correct position of the given number (= % deviation). Only the correctly calculated addition and subtraction problems were included, but the

percentage of correctly solved addition or subtraction trials was also calculated.

Working Memory

Spatial working memory was assessed by the Block Suppression Test (Beblo et al., 2004). This test is based on the CORSI-Block Tapping test (Schellig, 1997) and requires the subject to reproduce every second block in a given sequence of touched cubes on a wooden board as the examiner demonstrated. While the sequences gradually increase in length, the number of cubes last tapped in two consecutively correct sequences is defined as the maximum spatial working memory span.

Priming Task

To assess the affective effects of primes on calculation an adapted version of the task developed by Rubinsten and Tannock (2010) was used. It included four different types of primes (words with either positive, negative or neutral affect and words related to mathematics) and single-digit arithmetic problems (additions and subtractions) served as targets. As illustrated in **Figure 1**, each trial consisted of a prime presented aurally via headphones followed by an arithmetical problem. Reaction time in milliseconds was measured by the computer from the target onset to the participant's response. Each participant underwent in total 80 trials.

The primes were comprised of 40 words, including 10 per affective dimension (e.g., sun as a positive affective word, wood as a neutral affective word, prison as a negative affective word and count as a mathematic related word). The words were selected from the "The Berlin Affective Word List Reloaded" (BAWL-R), which contains a large set of psycholinguistic indexes known to influence word processing, and also features ratings of the emotional arousal, emotional valence and imageability of each word (Vö et al., 2009). Since the word ratings of the BAWL-R were based on ratings from 200 adults, we selected 61

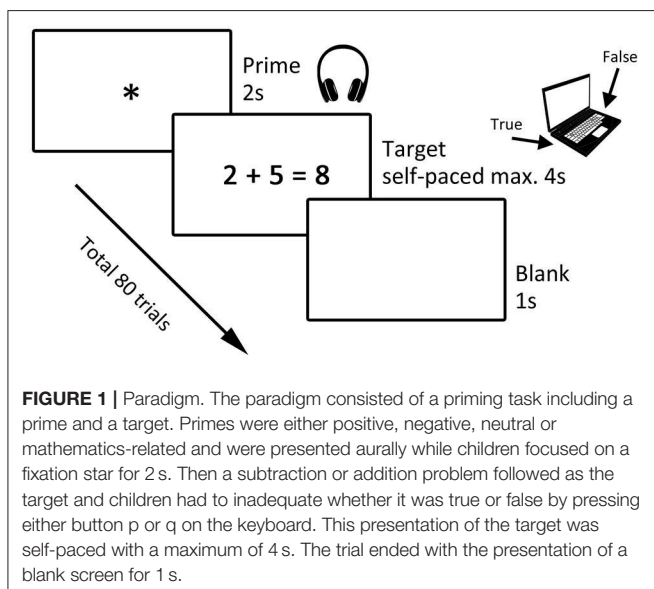
words, which were balanced for the number of letters, number of syllables, type of word (noun, verb, adjective), and emotional valence (positive, negative, neutral) or with mathematics related content and had them re-evaluated by a group of children. In total 123 children from the 2nd ($N = 29$), 3rd ($N = 30$), 4th ($N = 32$), and 5th ($N = 32$) grade rated these words. They were asked to indicate how they felt when they heard the word by marking a five-stepped smiley scale from happy to sad smileys, and to indicate if the word was related to mathematics, yes or no. Obtained ratings were analyzed for all children, as well as for each grade and for girls or boys separately. Consideration of these findings led to the final word list that was used in the present priming task. Please see Table S7 for a detailed description of the words.

The arithmetic problems were presented in the form " $a * b = c$," where a and b represent single digits from 1 to 9, $*$ represents an arithmetic operation (+ or -) and c represents the solution, which also consisted of only one digit (e.g., $2 + 1 = 3$ as a correct addition target). The type of arithmetic operation (each prime was once followed by an addition and once by a subtraction problem) and whether the problem was true or false were balanced between affective dimensions of primes and presented in a randomized order.

The present task differed from the original one of Rubinsten and Tannock (2010) in the following aspects: We have simplified the task by reducing the total number of trials from 160 to 80 by excluding multiplication and division arithmetic problems. Furthermore, the maximal response latency was extended from 3,000 to 4,000 ms. Moreover, we only included single digit solutions. Whereas Rubinsten and Tannock presented the primes visually in English, we presented the primes aurally in German. These changes in the priming task were conducted to adapt the paradigm to our younger cohort, consisting mostly of children in grades 2 or 3, whereas subjects in the study of Rubinsten and Tannock were in grades 4 or above.

Data Analyses

Data were analyzed by IBM SPSS Statistics Version 24 (IBM SPSS Statistics for Windows, 2016). Raw data of the priming task were extracted from E-Prime (E-Prime, 2002) and converted into SPSS. First, all behavioral data were tested for normality by the Kolmogorov-Smirnov test. If the data followed a normal distribution, groups were compared by parametric independent-sample t -tests. If data were not normally distributed, the nonparametric Mann-Whitney U -test for two independent samples was used. Nominal data (gender) was compared between control children and the DD group with a chi-squared test. P -values lower than 0.05 were considered statistically significant. To evaluate the effects of the priming task a general linear model analysis was conducted with RT as dependent variable. The 4 (type of prime: positive/negative/neutral/mathematics-related) $\times$ 2 (arithmetic operation: addition/subtraction) repeated measures ANCOVA defined type of prime and arithmetic operation as within-subject factors and group (CC/DD) as the between-subject factor. Since DD and CC groups differed in age, age was included as a covariate to exclude the possibility that group differences might be based on age differences. Regarding IQ,



it was expected that children with DD show a lower mean IQ compared to typically achieving peers as IQ measures are not fully independent from measures of math ability (Lambert and Spinath, 2018). In our analyses, we decided not to match groups on IQ, because one would have artificially influence the pattern of the normal population of DD or CC children. Moreover, IQ not to include as covariate in statistical analyses, which is in line with the suggestion of Dennis et al. (2009), who state that it is misguided and unjustified to attempt to control for IQ differences for cognitive outcome. However, we repeated the all tests with IQ and age as covariates showing that the main results did not change.

RESULTS

Demographic and Behavioral Data

Findings of demographic and behavioral data are summarized in **Table 1**. Statistical group comparisons indicated that children with DD were significantly older compared to control children, but groups did not differ in gender distribution. As expected, children with DD performed worse in all mathematical tests (addition and subtraction subtests of HRT, subtests of ZAREKI-R, Rechentest of BUEGA, and number line task). Even though all children had an IQ in the normal range, the IQ of the control group was significantly higher. In visuospatial working memory, no group differences were evident.

General findings regarding accuracy and reaction time between groups (CC/DD) or between conditions (addition/subtraction) are displayed in **Table 2**. For accuracy, the group comparison revealed significant differences between control and DD children in total accuracy and the accuracy of addition or subtraction problems separately, such that control children made significantly fewer errors. Furthermore, all children showed a higher accuracy for addition compared to subtraction (all children $N = 172$, $Z = -5.29$, $p < 0.001$; DD, $N = 76$, $Z = -3.56$, $p < 0.001$; CC, $N = 96$, $Z = -3.96$, $p < 0.001$). Results indicated that all participants were faster for addition problems compared to subtraction problems [all children $N = 172$, $t_{(170)} = -4.96$, $p < 0.001$; DD, $N = 76$, $t_{(75)} = -2.26$, $p < 0.05$; CC, $N = 96$, $t_{(94)} = -4.62$, $p < 0.001$], but no significant differences between control children or dyscalculic children were found (please see **Table 2**).

Affective Priming

To analyze potential *negative math priming effects*, a repeated measures ANCOVA was used. The type of prime (positive/negative/neutral/mathematics-related) and arithmetic operation (addition/subtraction) were defined as within-subject factors, group (CC/DD) was defined as the between-subject factor and age as a confounding variable. The analysis revealed no significant main effects for the type of prime [$F_{(3, 157)} = 1.087$, n.s.] or arithmetic operation [$F_{(1, 159)} = 0.149$, n.s.]. However, the interaction between type of prime and group reached significance [$F_{(3, 157)} = 3.762$, $p = 0.012$, $\eta^2 = 0.067$]. This interaction was further analyzed with *post-hoc* tests separated by group. As displayed in **Figure 2** and **Table 3**, in the *control children*, response latencies were significantly shorter for positive,

negative and neutral primes compared to mathematics-related primes. In the *dyscalculic children*, response latencies were significantly shorter for neutral than for positive, negative or mathematics-related primes. Hence, neither control children nor children with DD revealed a *negative math priming effect* as reported by Rubinsten and Tannock (2010).

Although we found no main effect of arithmetic operation in the present study, we also analyzed the effects of primes separately for addition and subtraction problems to allow direct comparison to the results of Rubinsten and Tannock (2010); please see Supplementary Material 1. Affective priming split by arithmetic operations, including Table S1). Similar to our main findings, no effects of prime were found. Regarding the arithmetic operation, reaction times (= *dependent variable*) were further contrasted for the groups depending on the type of prime. While in the *control group* significantly shorter response latencies were found for addition than for subtraction across all type of primes, in the *DD group* no significant differences were evident. Please see Supplementary Material 2. Differences between arithmetic operations split by primes, including Table S2).

The entire analysis was also performed after only including children who performed above chance level (mean accuracy $\geq 50\%$). Results of the repeated measures ANCOVA indicated no significant main effects or interactions. However, it is noteworthy that the interaction between type of prime and group missed significance only narrowly [$F_{(3, 142)} = 2.654$, $p = 0.051$, $\eta^2 = 0.053$]. For a detailed description of the demographic and behavioral data for this subgroup of children, please see Supplementary Material 3. Data analyses for accuracy levels above chance, including Table S3.

Gender Differences

In general, boys and girls tend to have different attitudes to mathematics such that girls express more concern about their mathematical performance (reviewed by Johns et al., 2005). Previous reports that girls show more mathematics anxiety (reviewed by Dowker et al., 2016), and that this math anxiety had an effect on mathematical performance lead us to the analyses of possible gender differences. Moreover and especially with regard to the affective priming task, it is important to note that gender differences have been reported in emotion processing too (reviewed by Hamann, 2004). Hence, for a more detailed understanding of possible gender differences in priming effects, analyses were performed after splitting the control and DD children into male and female subgroups. All results are presented first for the control and then for the DD children. For demographic and behavioral data please see Supplementary Material 4. Gender differences, including Tables S4, S5. In sum, control boys performed significantly better in arithmetic fluency, whereas this pattern was reversed in the DD children, and DD girls additionally performed better in working memory.

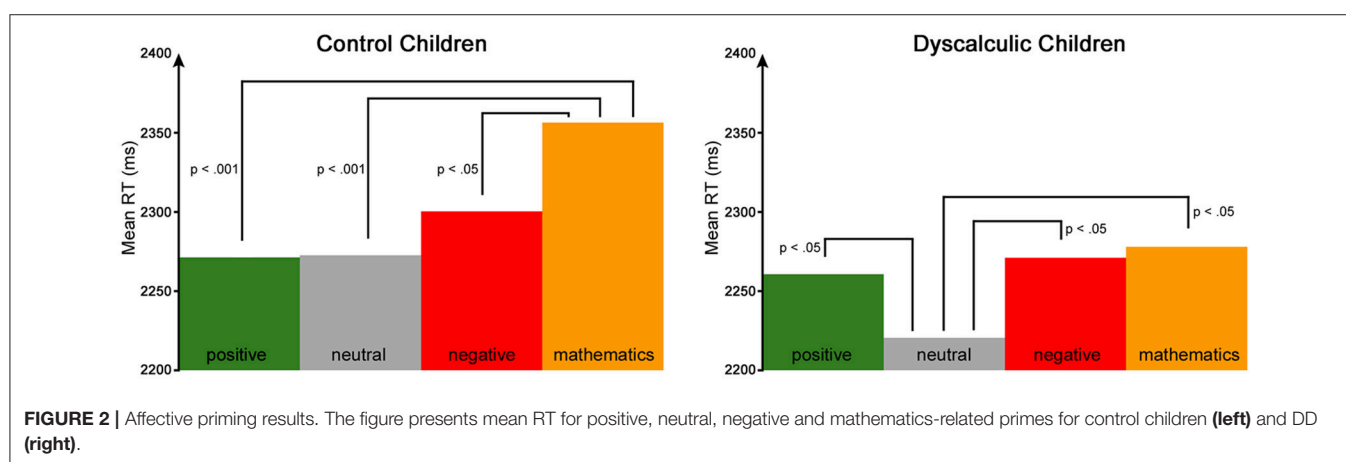
Boys and girls were then analyzed for priming effects. A repeated measures ANOVA was used in which type of prime (positive/negative/neutral/mathematics-related) and arithmetic operation (addition/subtraction) were defined as within-subject factors and gender as between-subject factor.

TABLE 2 | General findings on accuracy and reaction time.

	Total	DD	CC	Statistics between groups
Subjects (N)	172	76	96	
Accuracy (%) ^a <i>M (SD)</i>	72.71 (18.10)	66.33 (18.19)	77.76 (16.43)	$U = 2236, z = -4.36, p < 0.001$
Accuracy addition (%) ^a <i>M (SD)</i>	75.23 (19.06)	69.47 (19.62)	79.79 (17.40)	$U = 2446.5, z = -3.71, p < 0.001$
Accuracy subtraction (%) ^a <i>M (SD)</i>	70.23 (19.00)	63.29 (18.95)	75.72 (17.24)	$U = 2180, z = -4.53, p < 0.001$
Reaction time ^b (ms) <i>M (SD)</i>	2278.34 (492.26), <i>N</i> = 169	2239.94 (544.05), <i>N</i> = 75	2308.97 (447.30), <i>N</i> = 94	$t(142) = 0.89, n.s.$
Reaction time additions ^b (ms) <i>M (SD)</i>	2200.77 (412.23), <i>N</i> = 172	2225.68 (420.53), <i>N</i> = 76	2181.05 (406.67), <i>N</i> = 96	$t(170) = -0.70, n.s.$
Reaction time subtractions ^b (ms) <i>M (SD)</i>	2287.24 (393.14), <i>N</i> = 171	2282.29 (384.45), <i>N</i> = 76	2291.19 (401.95), <i>N</i> = 95	$t(169) = 0.15, n.s.$

^aMean accuracy of all 80 trials in %, or respectively of 40 trials for addition or subtraction.

^bMean reaction time including only correct trials within the range of ± 2.5 SD from the conditional mean latency in milliseconds (ms), for all, or only addition or subtraction problems.

**TABLE 3 |** Reaction times (RT) as a function of the prime valence in control children (CC) and in children with developmental dyscalculia (DD).

Prime valence	CC						DD					
	RT difference (ms)			df	t	p	RT difference (ms)			df	t	p
	N	Mean	SD				N	Mean	SD			
Pos—Neg	93	−29.06	248.82	92	−1.13	n.s.	74	9.70	331.84	73	0.25	n.s.
Pos—Neutr	93	12.73	223.59	92	0.55	n.s.	73	83.41	270.05	72	2.64	<0.05
Pos—Math	93	−88.65	228.52	92	−3.74	< 0.001	74	4.82	291.79	73	0.14	n.s.
Neg—Neutr	93	41.79	263.03	92	1.53	n.s.	72	71.51	298.82	71	2.03	<0.05
Neg—Math	93	−59.58	257.92	92	−2.23	<0.05	73	−2.61	290.75	72	−0.08	n.s.
Neutr—Math	94	−94.35	250.39	93	−3.65	<0.001	72	−78.72	287.65	71	−2.32	<0.05

In *control children*, significant main effects for type of prime [$F_{(3, 88)} = 7.977, p < 0.001, \eta^2 = 0.214$] and arithmetic operation [$F_{(1, 90)} = 27.888, p < 0.001, \eta^2 = 0.237$] were found, but no gender effects were observed. For *post-hoc t*-tests regarding prime or operation please see **Tables 2, 3**. In *dyscalculic children*, the repeated measures ANOVA revealed significant main effects for type of prime [$F_{(3, 66)} = 6.225, p = 0.001, \eta^2 = 0.221$] and operation [$F_{(1, 68)} = 5.192, p = 0.026, \eta^2 = 0.071$], but again no gender effects. Results of *post-hoc* tests for the effect of prime and operation are summarized in **Tables 2, 3**.

Taken together, no gender effects were evident in DD or control children and hence no gender-dependent *negative math priming effects* were found.

Mathematics Anxiety

The analyses were also repeated after taking into account the children's level of mathematics anxiety, quantified through the direct, explicit measure of the Math-Anxiety-Interview (please see section Cognitive Assessments). This complements the implicit measure of mathematics anxiety through the priming

task. The math anxiety interview was performed because it is known that mathematics anxiety has an impact on mathematical performance (reviewed by Wu et al., 2014). Hence, the performance in the task may not solely be influenced by the affective priming, but also by the extent of children's mathematics anxiety. Two different analyses with regard to mathematics anxiety were performed: firstly, a bivariate correlation was calculated between the level of mathematics anxiety (MAI) and the difference in reaction times between mathematics-related and negative affective primes. This was in order to test whether there exists a relationship between math anxiety and a possible *negative math priming effect*. As both variables were not normally distributed, a Spearman Correlation was carried out, which, however, was not significant ($r_s = -0.057$, $N = 172$, n.s.).

Secondly, to further analyse potential effects of mathematics anxiety, the sample was split according to different levels of mathematics anxiety (low vs. high). Hence, two new groups were formed by selecting the 25% of children with the lowest or highest MAI values, respectively. The lowest 25% of children had a MAI-value of 0-1 and the highest 25% of children had a MAI-value of 5.9-9.75. In order to avoid having to select between children with the same MAI-value, all children with the respective threshold value were included. That is the reason why samples are not perfectly equal in size, resulting in a total number of $N = 46$ in the low math anxious group and $N = 54$ in the high math anxious group. Demographic and behavioral data is included in the Supplementary Material 5. Mathematics anxiety, Table S6. In summary, both groups were matched for age and gender, but the high anxious subgroup included more DD children (72% vs. 26%). Accordingly, the high anxious group performed worse in different mathematical and general cognitive tasks.

The two groups were then further analyzed for priming effects of reaction time as the dependent variable. A repeated measures ANCOVA was used with type of prime (positive/negative/neutral/mathematics-related) and arithmetic operation (addition/subtraction) as within-subject factors and low vs. high mathematics anxiety as between-subject factor. Group (CC/DD) was included as covariate to control for unequal distribution of DD and control children in low vs. high mathematics anxiety subgroups. The main effect of prime [$F_{(3, 88)} = 3.695$, $p = 0.015$, $\eta^2 = 0.112$] and the interaction between prime and group (low vs. high anxious) reached significance [$F_{(3, 88)} = 3.389$, $p = 0.022$, $\eta^2 = 0.104$]. Hence, the interaction was further investigated separately for the low and high mathematics anxiety group.

As illustrated in Table 4, *post-hoc* tests showed that in the *low anxious children*, response latencies were significantly shorter for positive than for mathematics-related affective primes. Furthermore, response latencies were significantly shorter for neutral than for negative or math affective primes. In *high anxious children*, no differences in response latencies to the different primes were observed. In sum, no *negative math priming effects* were evident for low or high anxious children. These findings are comparable with the results when groups were split by DD and CC, where the results of the low anxious children reflect findings of CC group.

In addition to possible effects of math anxiety on priming, we further investigated general characteristics of math anxiety which were explicitly evaluated by the MAI. First, as listed in Table 1, children with DD suffer more often from math anxiety. Second, no gender differences were evident in CC (see Table S4) or in DD (see Table S5). Third, Pearson correlation between math anxiety and age revealed no relation between both measures when including all subjects ($r = 0.055$, $N = 172$, n.s.), however, in DD children a decrease of math anxiety over development was found ($r = -0.274$, $N = 76$, $p < 0.05$), but not in CC ($r = 0.030$, $N = 96$, n.s.). Finally, the relation between math anxiety and behavioral measures was further investigated by Pearson correlations (please see Table 5). Including all children, results indicated a significant relation between math anxiety and IQ, mathematical performance, arithmetic fluency, addition, subtraction, number line performance, as well as working memory. In DD children, math anxiety was significantly related to mathematical performance, arithmetic fluency, addition, and subtraction. In CC, math anxiety correlated significantly with mathematical performance, arithmetic fluency, addition, subtraction, and number line performance. All these relations demonstrate that higher levels of math anxiety were associated with worse performance.

DISCUSSION

Mathematics is often associated with negative attitudes and emotions in children and adolescents. However, little is known about the interactions between mathematical performance and negative emotions. Hence, this research gap was addressed by the present study. The aim was to elucidate the link between mathematics anxiety, negative emotions, low performance and deficiencies in mathematics abilities such as in children with DD.

To approach this question, an arithmetic-affective priming task was used in which the influence of a prime stimulus (positive/negative/neutral or mathematics-related) on a simple arithmetic operation (addition/subtraction) was tested in 172 children between 7.3 and 11.3 years of age. Approximately half of the children were diagnosed with DD.

Findings revealed, in line with our expectations, that all children were faster and made less errors for addition problems compared to subtraction. This is because subtractions often need more decomposing into smaller sub-parts and moreover, compared to additions, they are not commutative (e.g., $2 + 3 \neq 3 - 2$) (reviewed by Peters and De Smedt, 2017).

DD children performed worse in all mathematical tests and showed higher levels of mathematics anxiety, which remained significant when controlling for age, since the CC were slightly younger than the DD children. This is important to note, as it has been often been claimed, but so far only few studies examined and corroborated the increased levels of mathematics anxiety in subjects with disabilities in mathematics (Wu et al., 2014). Moreover, children with DD often suffer from additional psychiatric disorders like general anxieties (reviewed by Dowker et al., 2016).

TABLE 4 | *Post-hoc* tests for type of prime (prime valence) in low vs. high mathematics anxiety (MAI) children.

Prime valence	Low MAI group						High MAI group					
	RT difference (ms)			df	t	p	RT difference (ms)			df	t	p
	N	Mean	SD				N	Mean	SD			
Pos—Neg	44	−66.38	223.49	43	−1.97	n.s.	51	−5.09	342.77	50	−0.11	n.s.
Pos—Neutr	44	10.94	206.48	43	0.35	n.s.	50	40.32	291.76	49	0.98	n.s.
Pos—Math	44	−104.90	251.27	43	−2.77	< 0.05	52	−20.08	330.45	51	−0.44	n.s.
Neg—Neutr	44	77.32	222.41	43	2.31	< 0.05	49	41.90	296.25	48	0.99	n.s.
Neg—Math	44	−38.52	303.73	43	−0.84	n.s.	51	−24.84	302.90	50	−0.59	n.s.
Neutr—Math	44	−115.84	233.90	43	−3.29	< 0.05	51	−60.80	320.37	50	−1.36	n.s.

TABLE 5 | Pearson's correlation between math anxiety and behavioral measures.

	Math anxiety (intensity) ^a		
	Total	DD	CC
Intelligence (IQ) ^b	<i>r</i> = −0.329 <i>p</i> < 0.001 <i>N</i> = 172	<i>r</i> = −0.158 n.s. <i>N</i> = 76	<i>r</i> = −0.150 n.s. <i>N</i> = 96
Mathematical performance (T) ^c	<i>r</i> = −0.508 <i>p</i> < 0.001 <i>N</i> = 172	<i>r</i> = −0.340 <i>p</i> < 0.01 <i>N</i> = 76	<i>r</i> = −0.368 <i>p</i> < 0.001 <i>N</i> = 96
Arithmetic fluency (T) ^d	<i>r</i> = −0.523 <i>p</i> < 0.001 <i>N</i> = 172	<i>r</i> = −0.344 <i>p</i> < 0.01 <i>N</i> = 76	<i>r</i> = −0.395 <i>p</i> < 0.001 <i>N</i> = 96
Addition (% correct) ^e	<i>r</i> = −0.531 <i>p</i> < 0.001 <i>N</i> = 160	<i>r</i> = −0.531 <i>p</i> < 0.001 <i>N</i> = 71	<i>r</i> = −0.267 <i>p</i> < 0.05 <i>N</i> = 89
Subtraction (% correct) ^e	<i>r</i> = −0.542 <i>p</i> < 0.001 <i>N</i> = 160	<i>r</i> = −0.456 <i>p</i> < 0.001 <i>N</i> = 71	<i>r</i> = −0.393 <i>p</i> < 0.001 <i>N</i> = 89
Number line (% deviation) ^e	<i>r</i> = 0.313 <i>p</i> < 0.001 <i>N</i> = 160	<i>r</i> = 0.196 n.s. <i>N</i> = 71	<i>r</i> = 0.230 <i>p</i> < 0.05 <i>N</i> = 89
Number line (T) ^f	<i>r</i> = −0.284 <i>p</i> < 0.001 <i>N</i> = 172	<i>r</i> = −0.071 n.s. <i>N</i> = 76	<i>r</i> = −0.110 n.s. <i>N</i> = 96
Working memory (items) ^g	<i>r</i> = −0.218 <i>p</i> < 0.01 <i>N</i> = 169	<i>r</i> = −0.187 n.s. <i>N</i> = 75	<i>r</i> = −0.163 n.s. <i>N</i> = 94

^aMean intensity of math anxiety assessed by the math anxiety interview (MAI); 0 = no math anxiety, 10 = very high math anxiety.

^bMean IQ based on 4 subtests [verbal IQ and matrices test of BUEGA, block design and similarities subtest of WISC-IV (*n* = 153)], or based on 6 subtests of the WISC-IV [block design, similarities, digit span, picture concepts, vocabulary, arithmetic (*n* = 19)].

^cMean mathematical performance based on 4 subtests (addition and subtraction of HRT, Zahlenstrahl II of ZAREKI-R, Rechentest of BUEGA) (*n* = 153), or based on addition and subtraction of HRT and ZAREKI-R (*n* = 19) in *T*-values.

^dMean of addition and subtraction of the HRT in *T*-values.

^eBased on number line task. The percentage of correctly solved addition or subtraction problems, and the percentage of the deviation between the exact location on the number line and the marked location of the child.

^fBased on subtests number line I and II of ZAREKI-R in *T*-values.

^gBased on maximum number of correctly recalled items of the Corsi-Suppression test.

Regarding gender differences, our results indicated that girls and boys of the CC or DD group experienced math anxiety equally often. This is a promising result regarding the widely

discussed stereotype that females are expected to be worse in math related topics and that females experience more math anxiety. Our findings are consistent with research indicating that countries providing equal education for females and males show little or no gender differences in mathematical performance (Spelke, 2005; Kohn et al., 2013). The reason why no gender differences in math anxiety were evident might be due to increasing evidence that such gender differences only develop at adolescence as consequence of societal exposure to gender stereotypes (e.g., Johns et al., 2005), or female teachers who experience math anxiety themselves (Beilock et al., 2010). In contrast, several studies report that younger children in primary school do not exhibit gender differences in math anxiety (e.g., Dowker et al., 2012; Harari et al., 2013). Our findings are consistent with these reports since the children in our study were in primary school.

In general, studies suggest that math anxiety increases with age (reviewed by Dowker et al., 2016). The present study rather suggests that math anxiety is already present in 8-year old children, which is consistent with reports suggesting that math anxiety can be detected in the earliest stages of formal math learning in school (Wu et al., 2012; Sorvo et al., 2017). In addition, math anxiety in children with DD even seemed to decrease over development, which might be a positive effect of increased care, but would need specific investigation.

A large body of evidence confirms that math anxiety severely interferes with math learning and performance, both because math anxious people are more likely to avoid mathematical activities and because math anxiety usurps working memory resources (reviewed by Dowker et al., 2016). Similarly, the present findings also revealed that children with increased math anxiety performed worse in math related topics (mathematical performance, arithmetic fluency, addition, subtraction, number line performance), as well as, working memory. Since the majority of our children with math anxiety belonged to the DD group, the present findings are in line with the notion that DD children show specific deficits in visuospatial working memory, but it is not possible to further differentiate between effects of math anxiety or DD on different working memory profiles (Mammarella et al., 2015). Moreover, a significant relationship between math anxiety and IQ was found. In terms of domain-general abilities, it has been suggested that poor

intellectual conditions (e.g., poor abstract thinking or poor visuospatial skills) may contribute to the development of math anxiety (reviewed by Suárez-Pellicioni et al., 2016). However, the relation between math anxiety and IQ might also be a result of the group composition in the present study and can be rather attributed to the relation between math anxiety and math performance, since within groups (DD or CC) no relation between math anxiety and IQ was found.

With regard to priming effects however, the present data did not corroborate results reported by Rubinsten and Tannock (2010). Since no differences were found in reaction times to positive or negative affective primes, no standard affective priming effects were evident in control or DD children. Furthermore, no *negative math priming effect* was found in DD children. Thus, mathematics-related primes did not act affectively related to targets. In the study by Sarkar et al. (2014), the priming paradigm from Rubinsten and Tannock (2010) was adopted, however, no mathematics-related words were included and hence no direct comparison is possible between their results and our main finding regarding the *negative math priming effect*. Nevertheless, no significant effects of valence (positive or negative primes) were found, which is in line with our finding of absent standard priming effects.

In our analysis, no main effects of group, type of prime or arithmetic operation were evident. However, we found a significant interaction between type of prime and group, indicating that the reaction to positive, negative, neutral and mathematics-related primes is significantly different between DD and control children.

Overall, our major result is that we did not find the *negative math priming effect* observed by Rubinsten and Tannock (2010). In contrast, our analysis showed that control children's responses were significantly faster after the presentation of a positive, neutral or negative affective prime compared to a mathematics-related one. DD children were significantly faster after a neutral affective prime compared to a positive, negative or a mathematics-related one. This implies that control children's performance was inhibited by mathematics-related primes but was not affected differently by positive, neutral or negative primes. In contrast, DD children's performance was facilitated if the presented prime was neutral. Taken together, mathematics-related primes significantly interfered with processing the target in control children, but not in the DD group.

In conclusion, control children showed better performance if the prime had either an affective valence (positive or negative) or no valence (neutral), but not if there was a relation to mathematics. Thus, control children's processing seemed to be inhibited by mathematics-related primes. This might be a consequence of incongruent prime-target pairs. For example, if the mathematics-related prime word was "minus" and the subsequent target was an addition, this conflict might have disturbed processing in control children. It is implied that even if the prime word is instructed to be ignored, it automatically interferes with processing. This argument is supported by the Stroop effect, in which an interference is shown when processing an incongruent condition of font color and word meaning which results in slower reaction times (Stroop, 1935). Similar effects

have been reported in the context of mathematics, hence called the numerical Stroop effect, in which the physical size of the presented number interferes with the judgment of actual number size (e.g., Kaufmann et al., 2005). Accordingly, it seems plausible that mathematics-related prime words could interfere with subsequent calculation processes but only in control children. Similarly, these congruity effects have been reported in healthy participants, whereas they were not evident or much smaller in subjects with DD (Rubinsten and Henik, 2005). As found in the present study, this lends further support to the notion that dyscalculic subjects, unlike typically developing children, fail to process the irrelevant dimension automatically. In contrast, DD children performed faster if the prime had no valence (neutral). This is in contrast to what is expected according to the results of affective priming paradigms. As mentioned, faster processing is hypothesized if the prime-target pair is of the same valence (Hermans et al., 1994), which can already be observed in children (Spence et al., 2006). Since we assume that mathematics is negatively associated in children with DD, processing was expected to be faster if the prime had a negative valence or was mathematics-related.

In sum, since reaction times to positive or negative affective primes did not differ, the present study could neither find standard affective priming effects in control nor in DD children. Importantly, reaction times to negative or mathematics-related primes did not differ either, which clearly shows that no *negative math priming effect* in DD children was evident. A closer look at Rubinsten and Tannock's study, which suggested a *negative math priming effect* in DD, revealed that they did not observe the *negative math priming effect* in subtraction and division either. While they did observe it in addition and multiplication, an opposite pattern was shown for subtraction and for division the authors reported even a more facilitating influence of mathematics-related primes compared to negative primes. The fact that we did not observe the *negative math priming effect* in the present study might therefore be due to pooling addition and subtraction during the task. Therefore, we further analyzed our data separately for addition and subtraction. In line with Rubinsten and Tannock (2010), we did not observe the *negative math priming effect* in subtraction trials. In contrast, the *negative math priming effect* was absent in addition in the present study too.

To further investigate the inconsistent findings in terms of the *negative math priming effect*, several additional analysis were conducted that considered accuracy levels, gender and mathematics anxiety. As the amount of errors differed substantially between Rubinsten and Tannock (2010) and the present study, we decided to re-analyze the data and apart from solely including correct responses, only children that performed above chance level were considered. Consistent with our main findings, even when considering accuracy levels, no *negative math priming effect* was observable. Hence, the absence of the *negative math priming effect* cannot be justified by overall lower accuracy levels of children in the present study.

Furthermore, we carried out an analysis including gender, as previous research assumed attitudes and levels of mathematics anxiety to be different in boys and girls (reviewed

by Dowker et al., 2016). Hence, these differences were considered to affect priming effects. However, no priming effects were evident in boys or in girls, which again strengthens our major finding that we did not observe a *negative math priming effect*. Thus, gender is not a driving force of effects detected by Rubinsten and Tannock (2010). Regarding gender, it is also of interest that in our sample boys and girls of both the control and the DD group did not differ in mathematics anxiety. This is noteworthy as the literature reported inconsistent gender differences. Furthermore, math anxiety is supposed to increase with age during childhood and adolescence (reviewed by Dowker et al., 2016), which is a possible reason why no differences were evident in the present sample represented by rather young children. Similarly, empirical data including that from younger children also reported no differences in the levels of mathematics anxiety between boys and girls, since these differences only develop in adolescence (Johns et al., 2005).

Lastly, different levels of mathematics anxiety were taken into account as explanatory factors. The rationale underlying this analysis is the assumed negative influence of mathematics anxiety on mathematical performance (reviewed by Dowker et al., 2016). Hence, we hypothesized that priming effects could appear when groups were separated by extreme levels of mathematics anxiety. Concretely, this examination included a direct measure of mathematics anxiety (assessed by the MAI Kohn et al., 2013). This explicit measure of mathematics anxiety is an interesting add-on to our examination. Compared to Rubinsten and Tannock's implicit measure of mathematics anxiety through the priming paradigm, we consider the MAI to provide a more reliable assessment of mathematics anxiety. However, the potential relation between mathematics anxiety and the *negative math priming effect* was not significant. Furthermore, the analysis split into low vs. high anxious children found no *negative math priming effect* neither in the group of low, nor in the group of highly math anxious individuals. Therefore, mathematics anxiety is not causative for the priming effects observed by Rubinsten and Tannock (2010), which further supports our main result that no *negative math priming effect* was evident. Furthermore, these findings additionally support the notion that there is no relation between explicitly quantified math anxiety and the proposed implicit measure of math anxiety by the priming task. This leads to the conclusion that the priming task probably does not reliably assess math anxiety.

Taken together, further analyses showed that the *negative math priming effect* is independent of accuracy levels, gender, and mathematics anxiety, as it could still not be replicated when considering these variables. However, one might argue that present findings point toward an affective priming effect (comparing effects of positive and negative primes) since both the low math anxious subgroup and the control group reacted faster after a positive prime than after a negative prime. In the high math anxious subgroup and the DD group, this was not the case. This might hint to a possible positive valence of calculation problems in low math anxious or control children and a negative valence of arithmetical problems in high math anxious or DD children. Similarly, results seem indicative of a negative math priming effect (comparing effects of positive

and mathematical primes) as both low math anxious individuals and control children are faster after positive than mathematical primes, whereas this difference is not as big in highly math anxious children and appears to be opposite in DD children. However, since none of these differences reached significance, we cannot confirm an affective priming or a negative affective priming effect in the present study. Nonetheless, differences to Rubinsten and Henik (2005) must also be taken into account. For instance, in their study children were older than the ones included in the present study. As it is assumed that mathematics anxiety increases with age in childhood, the *negative math priming effect* might also be age dependent. Although we included age as a confounding variable in our analyses, age differences are still relevant for consideration of the *negative math priming effect*. In addition, the sample sizes differed as well. Rubinsten and Tannock (2010) included 23 children, whereas we considered data of 172 children. Our large sample size positively influences statistical power and the reliability of our results.

Importantly, the presentation of primes was different between both studies since Rubinsten and Tannock (2010) presented them visually whereas in our study the presentation was aural. This adjustment was made because our children were younger and hence potential difficulties in reading could be excluded through the aural presentation of primes. However, since the above mentioned congruity effects and the Stroop effect were both validated when visual primes were used, the aural presentation in our study might have affected priming effects. With regard to the literature, similar mechanisms have been shown to operate in visual and aural priming, but several differences in the priming effect have also been reported between the two modalities (Holcomb and Neville, 1990). In Holcomb and Neville's study, priming effects were larger in the aural condition but mean reaction times were slower in the aural than in the visual modality. Regarding the different presentation modalities between both studies, the interval between the onset of the prime word and the onset of the target (SOA), as well as the interval between the offset of the word and the onset of the target (ISI) also differed. Whereas in the present study the mean SOA was around 1 s and the ISI ranged from 0 to 567 ms, Rubinsten & Tannock used a shorter SOA of 250 ms and ISI of 0 ms. Klauer (1997) reported in his review that some priming experiments failed to obtain priming effects when using longer SOAs. This has also to be considered in the interpretation of the present findings. However, several studies presenting spoken prime words used comparable timing as in the present study and reported priming effects (Holcomb and Neville (1990) SOA = 1,420–1,850 ms, ISI = 1,150 ms; Voyer and Myles (2017) SOA = 800–1,000 ms, ISI = 50–250 ms; Kim and Sumner (2017) SOA = not indicated, ISI = 100 ms; Bacovcin et al. (2017) SOA = not indicated, ISI = 400–600 ms). Holcomb and Neville (1990) even reported stronger priming effects for auditory (SOA 1,420–1,850 ms) than visual prime words (SOA 1,550 ms), which had even longer SOA. Therefore, the longer SOA in the present study are unlikely to be the reason for absent priming effects, but should be further investigated.

Due to the younger age of the participants, we also simplified the task by excluding multiplications and divisions, which

resulted in fewer trials (80 compared to 160). Nevertheless, this should not be associated with insufficient power because in contrast to the lesser number of trials per person, we included many more participants in total.

A discrepancy that might have influenced the priming effects is the choice of language for the primes since Rubinsten and Tannock's study (Rubinsten and Tannock, 2010) was conducted in English and the present one in German (for a discussion of this effect for English and Hebrew see also Rubinsten et al., 2012). Certain characteristics of the languages, such as word length, frequency or part of speech (e.g., verbs, nouns or adjectives), may have influenced the priming effects. Moreover, priming words also differed between the studies. In the present study, the words for each category were carefully selected and evaluated in a pilot study with children to confirm that the chosen German words were affectively loaded either positive, negative, or neutral or clearly related to mathematics for children in the age range of the current study. Hence, language and words might influence the priming effects but as priming words were validated and pretested for valence, the language and selection of words should not be the sole cause of the inconsistent findings between the studies. Nevertheless, this issue needs further investigation.

To the best of our knowledge, apart from Rubinsten and Tannock (2010), the only other study that used an affective priming task and included mathematics-related words was conducted by Rubinsten et al. (2012). In spite of standard affective priming effects being observed by that study, no significant effects were found for mathematics-related primes. Thus, in line with the present finding, no *negative math priming effect* was evident.

In summary, the present study observed no priming effects and particularly not the *negative math priming effect* of concern in children with or without DD. However, we found that control children were significantly slower after the presentation of a mathematics-related prime compared to the positive / negative / neutral primes. In contrast, dyscalculic children were slower after a positive, negative or mathematics-related prime compared to a neutral prime. The present findings indicate that an affective math priming task, which is supposed to test the relation between emotions and mathematical performance in an implicit manner, might not be an ideal way to assess mathematics anxiety in the assessed age group. Rather, it might be more reliable to assess mathematics anxiety with explicit measures such as questionnaires.

A more detailed knowledge of the constructs is critical since "an understanding of the effects of math anxiety is fundamental

to understand the human cognitive apparatus in numerical abilities," as pointed out by Rubinsten and Tannock (2010, p. 10). Future research should also address the direction of causation between mathematical performance, mathematics anxiety and negative emotions. Because of the relatively high occurrence of math anxiety, an improved understanding of how these aspects relate to one another would enable interventions to be applied to improve performance in mathematics by reducing math anxiety and negative attitudes toward mathematics in school or math related situations in daily life.

AUTHOR CONTRIBUTIONS

KK: Conceptualization of the study idea and design, data collection, data analyses and interpretation and preparation and revision of this manuscript. IZ: Data analyses and interpretation and preparation of this manuscript. JK: Conceptualization of the study idea and design, data collection and revision of this manuscript. AW: Conceptualization of the study design and revision of this this manuscript. NP: Conceptualization of the study design and revision of this this manuscript. GE: Conceptualization of the study design and revision of this this manuscript. MvA: Conceptualization of the study design and revision of this this manuscript. All authors have contributed and have approved this final manuscript.

FUNDING

The present study was financially supported by a research grant of the German Federal Ministry of Education and Research (BMBF, number: 01GJ1011).

ACKNOWLEDGMENTS

Special thanks to Orly Rubinsten to provide us her affective priming paradigm and her support in study design and interpretation of our findings, to Nadine Häusler for her contribution in data entering and analyses, and to Ruth O'Gormann Tuura for proofreading.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.00263/full#supplementary-material>

REFERENCES

- APA (2013). *American Psychiatric Association: Diagnostisches und Statistisches Manual Psychischer Störungen - DSM-V*. Washington, DC: American Psychiatric Association.
- Bacovcin, H. A., Goodwin Davies, A., Wilder, R. J., and Embick, D. (2017). Auditory morphological processing: evidence from phonological priming. *Cognition* 164(Suppl. C), 102–106. doi: 10.1016/j.cognition.2017.03.011
- Beblo, T., Macek, C., Brinkers, I., Hartje, W., and Klaver, P. (2004). A new approach in clinical neuropsychology to the assessment of spatial working memory: the block suppression test. *J. Clin. Exp. Neuropsychol.* 26, 105–114. doi: 10.1076/jcen.26.1.105.23938
- Beilock, S. L., Gunderson, E. A., Ramirez, G., and Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1860–1863. doi: 10.1073/pnas.0910967107

- De Houwer, J., Teige-Mocigemba, S., Spruyt, A., and Moors, A. (2009). Implicit measures: a normative analysis and review. *Psychol. Bull.* 135, 347–368. doi: 10.1037/a0014211
- Dennis, M., Francis, D. J., Cirino, P. T., Schachar, R., Barnes, M. A., and Fletcher, J. M. (2009). Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *J. Int. Neuropsychol. Soc.* 15, 331–343. doi: 10.1017/S1355617709090481
- Dowker, A., Bennett, K., and Smith, L. (2012). Attitudes to mathematics in primary school children. *Child Dev. Res.* 8:124939. doi: 10.1155/2012/124939
- Dowker, A., Sarkar, A., and Looi, C. Y. (2016). Mathematics anxiety: what have we learned in 60 years? *Front. Psychol.* 7:508. doi: 10.3389/fpsyg.2016.00508
- E-Prime (2002). (Version 2.0). *Psychology Software Tools, Inc.* Available online at: <http://www.pstnet.com>
- Esser, G., Wyschkon, A., and Ballaschk, K. (2008). *BUEGA: Basisdiagnostik Umschriebener Entwicklungsstörungen im Grundschulalter*. Göttingen: Hogrefe.
- Haffner, J., Baro, K., Parzer, P., and Resch, F. (2005). *Heidelberger Rechentest: Erfassung Mathematischer Basiskompetenzen im Grundschulalter*. Göttingen: Hogrefe.
- Hamann, S. (2004). Individual differences in emotion processing. *Curr. Opin. Neurobiol.* 14, 233–238. doi: 10.1016/j.conb.2004.03.010
- Harari, R. R., Vukovic, R. K., and Bailey, S. P. (2013). Mathematics anxiety in young children: an exploratory study. *J. Exp. Edu.* 81, 538–555. doi: 10.1080/00220973.2012.727888
- Hermans, D., Houwer, J. D., and Eelen, P. (1994). The affective priming effect: automatic activation of evaluative information in memory. *Cogn. Emot.* 8, 515–533. doi: 10.1080/02699939408408957
- Holcomb, P. J., and Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: a comparison using event-related brain potentials. *Lang. Cogn. Processes* 5, 281–312. doi: 10.1080/01690969008407065
- IBM SPSS Statistics for Windows (2016). (Version 24.0). Armonk, NY: IBM Corp.
- Johns, M., Schmader, T., and Martens, A. (2005). Knowing is half the battle: teaching stereotype threat as a means of improving women's math performance. *Psychol. Sci.* 16, 175–179. doi: 10.1111/j.0956-7976.2005.00799.x
- Kaufmann, L., Koppelstaetter, F., Delazer, M., Siedentopf, C., Rhomberg, P., Golaszewski, S., et al. (2005). Neural correlates of distance and congruity effects in a numerical Stroop task: an event-related fMRI study. *Neuroimage* 25, 888–898. doi: 10.1016/j.neuroimage.2004.12.041
- Kim, S. K., and Sumner, M. (2017). Beyond lexical meaning: the effect of emotional prosody on spoken word recognition. *J. Acoust. Soc. Am.* 142, EL49–EL55. doi: 10.1121/1.4991328
- Klauser, K. C. (1997). Affective priming. *Eur. Rev. Soc. Psychol.* 8, 67–103.
- Kohn, J., Richtmann, V., Rauscher, L., Kucian, K., Käser, T., Grond, U., et al. (2013). Das Mathematikangstinterview (MAI): erste psychometrische gütekriterien. *Lern. Lernstör.* 2, 177–189. doi: 10.1024/2235-0977/a000040
- Krause, S., Back, M. D., Egloff, B., and Schmukle, S. C. (2012). A new reliable and valid tool for measuring implicit self-esteem: the response-window affective priming task. *Eur. J. Psychol. Assess.* 28, 87–94. doi: 10.1027/1015-5759/a000095
- Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schonmann, C., Plangger, F., et al. (2011). Mental number line training in children with developmental dyscalculia. *Neuroimage* 57, 782–795. doi: 10.1016/j.neuroimage.2011.01.070
- Kucian, K., and von Aster, M. (2015). Developmental dyscalculia. *Eur. J. Pediatr.* 174, 1–13. doi: 10.1007/s00431-014-2455-7
- Lambert, K., and Spinath, B. (2018). Are WISC IQ scores in children with mathematical learning disabilities underestimated? The influence of a specialized intervention on test performance. *Res. Dev. Disabil.* 72(Suppl. C), 56–66. doi: 10.1016/j.ridd.2017.10.016
- Mammarella, I. C., Hill, F., Devine, A., Caviola, S., and Szucs, D. (2015). Math anxiety and developmental dyscalculia: a study on working memory processes. *J. Clin. Exp. Neuropsychol.* 37, 878–887. doi: 10.1080/13803395.2015.1066759
- OECD (2013). *PISA 2012 Results: Ready to Learn: Students' Engagement, Drive and Self-Beliefs (Volume III)*. PISA, OECD Publishing. doi: 10.1787/9789264201170-en
- Petermann, F., and Petermann, U. (2007). *Hamburg-Wechsler-Intelligenztest für Kinder IV (HAWIK-IV) [Wechsler Intelligence Scale for Children - Fourth Edition - German adapted version]*. Bern: Huber.
- Peters, L., and De Smedt, B. (2017). Arithmetic in the developing brain: a review of brain imaging studies. *Dev. Cogn. Neurosci.* doi: 10.1016/j.dcn.2017.05.002. [Epub ahead of print].
- Rubinsten, O., Bialik, N., and Solar, Y. (2012). Exploring the relationship between math anxiety and gender through implicit measurement. *Front. Hum. Neurosci.* 6:279. doi: 10.3389/fnhum.2012.00279
- Rubinsten, O., and Henik, A. (2005). Automatic activation of internal magnitudes: a study of developmental dyscalculia. *Neuropsychology* 19, 641–648. doi: 10.1037/0894-4105.19.5.641
- Rubinsten, O., and Tannock, R. (2010). Mathematics anxiety in children with developmental dyscalculia. *Behav. Brain Funct.* 6:46. doi: 10.1186/1744-9081-6-46
- Sarkar, A., Dowker, A., and Cohen Kadosh, R. (2014). Cognitive enhancement or cognitive cost: trait-specific outcomes of brain stimulation in the case of mathematics anxiety. *J. Neurosci.* 34, 16605–16610. doi: 10.1523/jneurosci.3129-14.2014
- Schellig, D. (1997). *Block-Tapping-Test*. Lisse: Swets & Zeitlinger.
- Shalev, R. S., Auerbach, J., Manor, O., and Gross-Tsur, V. (2000). Developmental dyscalculia: prevalence and prognosis. *Eur. Child Adolesc. Psychiatry* 9(Suppl 2), S58–S64. doi: 10.1007/s007870070009
- Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., et al. (2017). Math anxiety and its relationship with basic arithmetic skills among primary school children. *Br. J. Educ. Psychol.* 87, 309–327. doi: 10.1111/bjep.12151
- Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: a critical review. *Am. Psychol.* 60, 950–958. doi: 10.1037/0003-066X.60.9.950
- Spence, S., Lipp, O. V., Liberman, L., and March, S. (2006). Examination of emotional priming among children and young adolescents: developmental issues and its association with anxiety. *Aust. J. Psychol.* 58, 101–110. doi: 10.1080/00049530600730468
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662.
- Suárez-Pellicioni, M., Núñez-Peña, M. I., and Colomé, Á. (2016). Math anxiety: a review of its cognitive consequences, psychophysiological correlates, and brain bases. *Cogn. Affect. Behav. Neurosci.* 16, 3–22. doi: 10.3758/s13415-015-0370-7
- Vö, M. L. H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., and Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behav. Res. Methods* 41, 534–538. doi: 10.3758/BRM.41.2.534
- von Aster, M., Weinhold Zulauf, M., and Horn, R. (2006). *ZAREKI-R (Neuropsychological Test Battery for Number Processing and Calculation in Children), Revised Version*. Frankfurt: Harcourt Test Services.
- Voyer, D., and Myles, D. (2017). The effect of unimodal affective priming on dichotic emotion recognition. *Laterality*. doi: 10.1080/1357650X.2017.1404095. [Epub ahead of print].
- Werner, B., and Rothermund, K. (2013). Attention please: no affective priming effects in a valent/neutral-categorisation task. *Cogn. Emot.* 27, 119–132. doi: 10.1080/02699931.2012.711744
- WMA (2002). *The World Medical Association's Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects*. Washington, DC: WMA General Assembly. Available online at: <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>
- Wu, S. S., Barth, M., Amin, H., Malcarne, V., and Menon, V. (2012). Math anxiety in second and third graders and its relation to mathematics achievement. *Front. Psychol.* 3:162. doi: 10.3389/fpsyg.2012.00162
- Wu, S. S., Willcutt, E. G., Escovar, E., and Menon, V. (2014). Mathematics achievement and anxiety and their relation to internalizing and externalizing behaviors. *J. Learn. Disabil.* 47, 503–514. doi: 10.1177/0022219412473154

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

Copyright © 2018 Kucian, Zuber, Kohn, Poltz, Wyschkon, Esser and von Aster. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Individual Differences in Fourth-Grade Math Achievement in Chinese and English

Nicola A. McClung^{1*} and Diana J. Arya²

¹ Learning & Instruction, University of San Francisco, San Francisco, CA, United States, ² Department of Education, Gevirtz Graduate School of Education, University of California, Santa Barbara, Santa Barbara, CA, United States

OPEN ACCESS

Edited by:

Bert De Smedt,
KU Leuven, Belgium

Reviewed by:

Bieke De Fraine,
KU Leuven, Belgium
Christine Schiltz,
University of Luxembourg,
Luxembourg

*Correspondence:

Nicola A. McClung
namcclung@usfca.edu

Specialty section:

This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Education

Received: 20 January 2018

Accepted: 12 April 2018

Published: 04 May 2018

Citation:

McClung NA and Arya DJ (2018)
Individual Differences in Fourth-Grade
Math Achievement in Chinese and
English. *Front. Educ.* 3:29.
doi: 10.3389/feduc.2018.00029

Language has been widely acknowledged as a determining factor in mathematical achievement. Less understood, however, is the relationship between students' language and their performance on tests of mathematics when taking into consideration the presence of mathematical difficulties. We investigated the effects of two different language systems, Chinese and English, on the mathematical performance of fourth-grade (or age equivalent) students ($N = 23,220$) with varying levels of demonstrated mathematical and reading ability. For this investigation, we used a subset of the 2011 Progress in International Reading and Literacy Study (PIRLS) and Trends in International Mathematics and Science Study (TIMSS) from students who were tested in Chinese or English in nine countries. Findings from hierarchical linear modeling (HLM) analyses revealed that the main effect of language on mathematical performance remained significant once variables for mathematical ability were added to the model. Further, significant language-by-mathematical ability interactions were observed when controlling for country, gender, maternal education, and age. Thus, the effect of language on mathematical performance may be especially salient in the presence of mathematical difficulties. Implications of these findings include the need for further investigations of language and its effects on mathematical performance for Chinese- and English-speaking students in order to clarify how this relationship may vary within specific language populations.

Keywords: mathematics education, cross-linguistic, english, chinese, multilevel models, elementary education, individual differences

INTRODUCTION

Despite the assumed relative universality of mathematical knowledge and algorithmic processes, the effectual relationship between math performance and numerical language has been well established (Miller et al., 2005). Many investigations into the cross-national/linguistic differences in math performance focus on Chinese and English speaking populations, and there is considerable evidence that children in China, Taiwan, Singapore, and Hong Kong, historically and currently outperform students in Australia, Ireland, Canada, England, Scotland, and the U.S.; these consistent, and even dramatic, differences in math achievement (Peak, 1996, 1997; Mullis et al., 2016) have been attributed to a number of child and contextual factors, including maturation, caregiver values and beliefs, instructional method, and the degree of transparency of the number naming system.

Cross-linguistic researchers have hypothesized that less transparent number naming systems (i.e., languages in which there are a larger number of unique, irregular, or opaque words for numbers and mathematical concepts) may be more difficult to learn than more obvious number naming systems (i.e., language systems in which number names are more logically ordered to include names of earlier numbers and mathematical concepts are more readily understood from numerical language), and such differences have been found to have an effect on the mathematical performance of children (e.g., Miller et al., 1995, 2004; Göbel et al., 2014) and number processing of adults (Moeller et al., 2015). However, what is less understood is whether the linguistic influences on mathematical learning play out equally for all speakers of a language, or if numerical language is particularly influential on subsets of learners—such as students who have difficulty in math.

Within the parallel field of reading development, it is well understood that the processes required for reading are language specific (e.g., Seymour et al., 2003; Frost, 2005). For example, learning to read in *transparent* (more consistently spelled according to distinct sounds represented; e.g., Spanish, Finnish, Welsh) alphabetic writing systems develops more quickly than in *opaque* (less transparent; English, French, Portuguese) alphabetic orthographies. Furthermore, learning to read in a logographic writing system, such as Chinese, may be uniquely demanding, as the learner acquires symbol/sound relationships as well as memorizes thousands of written characters that directly correspond to meaning (Perfetti et al., 2005; Tan et al., 2005).

However, the effect of orthographic depth (which characterizes alphabetic languages) on reading development has also been found to have the most powerful effect on students who have the most difficulty learning to read; specifically, higher rates of reading impairment have been observed in students who speak orthographically opaque languages such as English compared to students who speak more transparent languages such as Spanish (Caravolas, 2005). This unique effect of language on lower-performing students' reading development is evidenced in a study by Hanley et al. (2004): 6- and 7-year-old Welsh-speaking children, who were learning to read in the highly predictable Welsh orthography, performed significantly better on reading measures than Welsh English-speaking children learning to read in English. However, after these students had reached their sixth year of formal instruction, while the majority of the English-speaking children had caught up to their Welsh-speaking counterparts on word reading (and even had significantly greater reading comprehension skills), the lowest performing 25% of English readers continued to perform significantly below the lowest performing 25% of Welsh readers on all measures of reading achievement.

Similarly, researchers have investigated the cognitive underpinning of reading difficulty in Chinese and compared results to those in alphabetic languages (Bolger et al., 2005). For example, Siok et al. (2004) found that reading impairment in Chinese was specific to the logographic nature of the writing system—pointing to the possibility that it is possible to have reading difficulty in Chinese but not in English, or vice versa,

depending on the individual's pattern of cognitive strengths and weaknesses that manifest according to the linguistic context.

Thus, while the association between domain-specific achievement and language has been observed for both math and reading, little is known about whether the challenges associated with learning math in the context of relatively opaque numerical language such as English (compared to Chinese) may be unique for the subset of students with the most difficulty learning math. Research in reading (e.g., Hanley et al., 2004) draws attention to the possibility that the linguistic influences on math competencies might be the most profound and long lasting for students with the lowest performance in math.

While we acknowledge the complex set of influences on math achievement, our focus here is on language. Specifically, we are interested in the possibility that Chinese and English numerical language may differentially affect students with the lowest demonstrated math ability. For this study, we aimed to further clarify differences in math performance that have been consistently observed across languages (e.g., Peak, 1996, 1997; Miller et al., 2005; Göbel et al., 2014). Specifically, we asked the following research questions:

- (1) What is the relationship between math performance and language for children who demonstrate high vs. low levels of mathematical skill?
- (2) What is the relationship between math performance and language for students who are dominant Chinese and English speakers?

We employed a subset of the 2011 Progress in International Reading and Literacy Study (PIRLS) and Trends in International Mathematics and Science Study (TIMSS) that included Chinese- or English-speaking students ($N = 23,220$) in order to investigate the relationships between language, math difficulty, and co-occurring reading and math difficulty at the fourth grade level. We posited that greater descriptive ambiguity in the relationship among mathematical concepts and the English words used to label them places additional demands on the learner and can therefore compromise or slow down fourth grade mathematical performance in English relative to Chinese. Furthermore, mathematical weaknesses, whether cognitive (Imbo et al., 2014) or due to environmental factors may be exacerbated when there is relatively less obviousness between mathematical concepts and the language used to represent them.

Learning to count base-10 Arabic numerals is a universal early math skill (Miller et al., 1995). However, the way in which labels or names map on to numerical items and the connotations implied to such items differ across languages (Hurford, 1975, 1987). For example, the word/character for “triangle” in Chinese is “三角,” which literally translates as “three cornered shape.” The first portion of this word (三) is the number 3, which is highly accessible to younger students due to the intuitive connections with the three lines represented. The English word “triangle” has a less transparent connection with the shape; one must understand that the morpheme “tri” indicates the meaning of three. Such linguistic differences may have a great impact on student learning of mathematical content like geometry (Miller et al., 2005).

Differences in Math Achievement in Chinese and English

Variation in Chinese and English Number Naming Systems

Although cross-linguistic differences in achievement have been attributed to differences in culture (Stevenson et al., 1986), curricula, and instruction (Stigler et al., 1982; Stevenson and Stigler, 1992) and language ideology (Arya et al., 2015), there is considerable evidence that variability in math performance in young children is, at least in part, associated with the variation in number naming systems. Such variation is thought to affect the relative ease and effectiveness with which children develop, access, store, and manipulate mathematical information (Miller et al., 1995; Imbo et al., 2014).

A comparison of the Chinese and English number naming systems highlights the characteristics of language that may affect the acquisition of math skills. Counting to 10 in both languages, for example, is similar in that the words used are unique to each number (Miller et al., 1995). However, after 10, the two languages differ in terms of the degree to which the names for larger numbers systematically include the same names used for the earlier numbers. Overall, the Chinese system is more morphographically obvious (as previously described) and involves less modification of the unit value number names (or fewer additional unique names) in larger numbers than the English system. For example, the Chinese names for numbers “11” and “12” are the equivalent to stating the name for “10” plus the name for the additional amount (ten-one; ten-two, etc.). This system is more obvious than the English system, which involves unique names (eleven, twelve, etc.). Thus, in Chinese, counting involves memorizing only the number names for 1 through 10 and then applying the base-10 rules to generate larger numbers (Okamoto, 2017). Furthermore, because counting and number representation are foundational to higher-level math skills, variability in the characteristics of number naming systems that affects these basic skills may have a long-term effect on achievement (Miller et al., 2004). In this study, we hypothesize that students with relatively weak or imprecise mathematical knowledge (compared to their peers who speak the same language) will face an additional challenge to mathematical learning when learning in the context of ambiguous or irregular numerical language (i.e., English compared to Chinese). Furthermore, we posit that specific weaknesses in mathematical understanding may be exacerbated by less obvious (i.e., less concretely descriptive) number naming systems within a given language context. As such, assessment items that feature words along with numbers to probe mathematical knowledge (including written directions and word problems) may present greater challenges when such inscriptions are less transparent.

Language Influences on Math Performance

Dramatic differences in math achievement in Chinese and English have been observed (e.g., Husen, 1967; Stevenson et al., 1986; Travers et al., 1987). For example, children in China have been shown to outperform U.S. children as early as

preschool, which has been attributed in part to the relative transparency of the Chinese number naming system (Miller et al., 1995). Indeed, students from Chinese-speaking countries (e.g., China, Taiwan, Singapore, and Hong Kong) continue to score consistently higher than students from English-speaking countries (e.g., Australia, the U.S., Ireland, Scotland, and England) on international measures of fourth-grade mathematics achievement (Peak, 1996; TIMSS, 2011); and such differences have been found to increase as students advance throughout schooling (Stevenson et al., 1998; OECD, 2010). However, it is important to acknowledge the well-documented differences in curricula, instructional approaches, parental support, language ideologies, and educational systems between Asian and English-speaking countries, which undoubtedly contribute to observed differences in achievement (e.g., Stigler et al., 2000; Hiebert et al., 2003, 2005; Arya et al., 2015). Language is certainly only one of many contributors to cross-national and cross-linguistic differences in math achievement.

Findings from several cross-linguistic studies suggest an association between number names and math development when comparing Chinese and English. For example, Miller and Stigler (1987) found differences in math skill between the two languages emerged prior to any formal schooling, thus potentially ruling out the possibility that cross-national differences could be attributed to variation in instructional method. Moreover, Miller et al. (2004) found in their longitudinal study that differences in skill acquisition between Chinese and English could be observed precisely at the point in development when children are first learning numbers between 10 and 20, the point from which the consistency of the two number naming systems differ according to naming obviousness. Miller and colleagues conducted monthly “counting task” sessions with preschool-aged children in China and the United States. Beginning at age two, children from both countries demonstrated equally slow and error prone performance, and that there were no language differences in the ability to count to 10 for three or four year olds. However, by age three, the Chinese group made rapid progress in learning to count accurately between 10 and 20 compared with the American children. Furthermore, by age four, after the Chinese children had learned to count past 40, they were more readily able than the American children to generalize from the consistent Chinese number naming rules to count to 100. These researchers also concluded that the ability to accurately and rapidly name numbers and count may support higher-level math skills.

Differences in arithmetical skills (Miura et al., 1988; Fuson and Kwon, 1992), such as borrowing and carrying and math processing (Moeller et al., 2015) have been observed in Chinese/English comparisons as well as across other languages. Furthermore, there is some indication that individual differences demonstrated within a given linguistic context must be considered when investigating students’ mathematical abilities. For example, Imbo et al. (2014) compared French and Dutch speaking children and found that both cognitive resources and language to played a role in number processing. Similarly, Miura (1987) and Miura and Okamoto (1989) showed that children who spoke relatively regular (e.g., Chinese and Japanese) vs. irregular (English and Swedish) number names developed different mental

constructions of numbers, and argue that Chinese-based number systems uniquely influence how children mentally represent numbers.

However, while many researchers exploring cross-linguistic influences on arithmetic performance have controlled for general cognitive abilities (e.g., Helmreich et al., 2011), to our knowledge little is known about whether students with particular cognitive profiles are more or less sensitive to the relative ambiguity in the math language environment. Thus, what continues to be less clear from all the research presented thus far is whether this effect of numerical language (specifically, in this case, Chinese versus English) is the same for all students or whether the characteristics of numerical language may be particularly advantageous (Chinese) or detrimental (English) for students who already have difficulties in the domain of mathematical learning. That is, do students with weak, or imprecise representations of number and/or mathematical concepts (for any reason—e.g., lack of exposure or practice, instructional method, cognitive impairment) encounter an additional and/or ongoing obstacle to math fluency when learning in less obvious numerical language?

Mathematics Difficulty

A considerable number of children have been found to have difficulty with number representation (i.e., number symbols, number names, and their corresponding magnitudes; Geary et al., 1999, 2004; Passolunghi and Siegel, 2004; Rousselle and Noël, 2007), the conceptual understanding of counting (Geary et al., 1992), counting speed (Passolunghi and Siegel, 2004), counting strategies (Goldman et al., 1988), monitoring the counting process (Jordan and Montani, 1997), and storing and retrieving number problems and solutions during mental arithmetic (Geary, 1993).

This research has typically compared the cognitive profiles and mathematics performance of three groups: (1) children with math difficulty alone (math only), (2) children with co-occurring math and reading difficulty (math/reading), and (3) control children. The goal in studying these subgroups has been to identify and describe “pure” math impairment, and to acknowledge and elucidate the challenges faced by those children who have difficulty with both math and reading. For example, in a longitudinal study, Jordan et al. (2003) compared these three subsets of children. They found that the math-only group demonstrated significantly slower and less accurate calculation strategies with difficulty drawing on numerical information from memory. While the math/reading group was observed to have similar problems, their demonstrated weaknesses were even more severe compared with the math-only and control groups, suggesting that in addition to arithmetical challenges, phonological weaknesses may also contribute to mathematics performance; thus, linguistic ability (i.e., phonological processing) may also play a significant role in math ability, both of which in turn may play a role in learning new and higher-level math skills.

Thus, it stands to reason that English-speaking children who demonstrate math difficulty might encounter additional detrimental effects (relative to Chinese-speaking children) of learning math via language that less transparently corresponds to

number names and symbols and the magnitudes and concepts they represent, which may lead to later weakness in engaging more complex calculations and problem-solving strategies.

The Current Study

We aimed to explore the relationship between math and reading ability and language by examining data at the fourth-grade level. Guided by previous research on math impairment (e.g., Swanson and Jerman, 2006) and cross-national (Peak, 1996, 1997) and cross-linguistic (Miller et al., 2005) differences in math performance, we selected variables from two large international databases, 2011 Progress in International Reading and Literacy Study (PIRLS) and Trends in International Mathematics and Science Study (TIMSS), to investigate our research questions about fourth-grade Chinese- and English-speaking children living in nine countries ($N = 23,220$). Specifically, we investigated the potential unique and lasting challenges of ambiguous numerical language on math learning for the lower performing students—by comparing Chinese and English results from the TIMSS Geometric Shapes and Measures and Data Display assessments.

METHODS

Sample and Data Sample

The sample includes 23,220 fourth-grade (or age equivalent 9.5–10.5) students from Australia, Taiwan, Hong Kong, Ireland, Malta, Northern Ireland, Qatar, Saudi Arabia, and Singapore who took part in 2011 PIRLS and TIMSS in Chinese or English. Students were included in this study if they took both assessments in either Chinese or English and spoke the language of the test at home (as indicated on a home survey). Thus, bilingual or multilingual students were dropped if they had a primary language other than the test language at home (Australia 6%, Taiwan 3%, Hong Kong 2%, Ireland 5%, Malta 5%, Northern Ireland, 1%, Qatar 44%, Saudi Arabia 65%, and Singapore 8%). Less than one percent of students who met our inclusion criteria were excluded because of missing data. To clarify, PIRLS, and TIMSS were not given in China, and students in the U.S. sat for either PIRLS or TIMSS, but not both assessments. As such these countries they were excluded from this study.

Data

The TIMSS and PIRLS assessments are conducted by the International Association for the Evaluation of Educational Achievement (IEA), and funded by the participating countries with support from the World Bank and the U.S. Department of Education's National Center for Educational Statistics (NCES; Martin and Mullis, 2012). Occurring every four (TIMSS) and five (PIRLS) years at the fourth-grade level (or its national equivalent), these assessment instruments are intended to provide internationally comparable information about mathematics, science, and reading literacy. In 2011, the TIMSS and PIRLS implementation came into alignment for the first time, and 34 countries took the opportunity to administer both TIMSS and PIRLS to the same students.

In using TIMSS and PIRLS datasets, our study involved only passive observation of publically available data, which did not contain identifying information, and thus ethics approval was not required per our institutional guidelines or national regulations.

Sampling Methodologies

All countries used a uniform sampling approach that followed international guidelines and specifications to ensure that differences in national achievement outcomes could not be attributed to the use of different sampling methodologies. Two-stage stratified sample designs were used, and probability samples were drawn from target populations (i.e., populations with the language as either English or Chinese) in each country (Mullis et al., 2009).

Participant Criteria

The TIMSS and PIRLS participants were representative samples of students in approximately their fourth year of formal schooling and who were between the ages of 9.5 and 10.5 who sat for both tests during the fall of 2011. Candidate participants for both studies are required to be able to follow basic instructions on the tests, and be able to read or speak the language of the test. Students with dyslexia and other learning disabilities were encouraged to participate in both PIRLS and TIMSS. The number of students excluded based on the above criteria did not exceed 5% in any country (Mullis et al., 2009).

Translation

In any cross-national study, it is critical that the measures are reliable and contain comparable information across languages. The development of TIMSS and PIRLS included exhaustive procedures to verify that the translation of the assessments corresponded to international standards, and to ensure equality across languages. Translation was provided for the test directions, passages, and items, student, home, and school questionnaires, directions for preparing and administering the assessment at schools, and scoring guides for students' open response questions (Mullis et al., 2009).

Math Achievement

In this study, math achievement was based on standardized performance ($M = 0$, $SD = 1$) on two of the three TIMSS content domains: Geometric Shapes and Measures (GSM) and Data Display (DD). In the GSM subsection, performance included the ability to measure and compare length, area, volume, and angle by drawing on knowledge about which units to use in each context. Students were required to approximate and estimate, and they used mathematical formulas to calculate the perimeter of rectangles and the volume of geometric figures. Data Display involved organizing, interpreting, and representing data. For example, students had to compare different types of data to make inferences, answer questions, and draw conclusions.

The development and validity check of the TIMSS achievement measures involved the use of item response theory (IRT), which enables the ability to analyze the relative level of difficulty of each individual item within a single measure and to use this information to determine the internal consistency of a given measure for the targeted domain of knowledge

(e.g., Geometric Shapes). TIMSS measures were developed in workshops within the representative countries by respective researchers and educators who reviewed the items and passages extensively. The TIMSS assessment in this study was comprised of two domains: Geometric Shapes and Measures (GSM) and Data Display (DD). Each of these cognitive domains captured a range of processes involved in math problem solving: Knowing, Applying, and Reasoning. The format of the TIMSS items was multiple-choice and constructed-response. Overall reliability of all math items were estimated within the range of $\alpha = 0.80$ – 0.89 . Reliability estimates for specific math subtests were not available.

Comparison Groups

Drawing on previous research (e.g., Swanson and Jerman, 2006), we compared the math performance (i.e., GSM and DD) of three groups of students: (1) students with math difficulty (MM) only, (2) students with both math and reading difficulty (MD/RD), and (3) students with average or above average math performance (not MD or MD/RD) in Chinese and English. In each language, we included approximately the same percentage of children in these groups. The specific grouping criteria are described below.

Mathematics Difficulty (MD)

Having difficulty in math (only) was determined by student performance on the Number content domain of the TIMSS assessment. This subsection measured number representation, knowledge of place value, and the relationship between numbers. Students demonstrated an understanding of and computational fluency in addition, subtraction, multiplication, and division. This subsection of the TIMSS for fourth grade is considered to be the most basic and foundational of all the subsections (cf., TIMSS, 2011) and is thus a useful (albeit limited) proxy for potential math difficulty. Further, using the Number domain subsection as an indicator of math difficulty aligns with previously described studies that documented the long-term effects of basic computational ability on the performance of more complex tasks (cf., Miller et al., 2004). Math difficulty was operationalized as being above the 10th percentile in reading (see below) but below the 10th percentile on the Number subsection within his or her language group. These criteria were within the range of scores used to operationalize mathematics difficulty in previous research (below the 48–8th percentile on various math measures; Swanson and Jerman, 2006). (The development of TIMSS is described in the section above.).

Co-occurring Math and Reading Difficulty (MD/RD)

Students with co-occurring MD/RD performed below the 10th percentile on the Number subsection of TIMSS and below 10th percentile within his or her language group on a relatively simple measure of reading achievement: the PIRLS' "Straightforward Processing" subsection. This scale measured the reader's ability to answer questions about information explicitly stated in the text, a skill that largely relies on efficient word recognition, which, in turn, is supported by phonological processing (e.g., Vellutino, 1979). Specifically, students had to read the text, access meaning on a basic level, and retrieve information contained directly in the text

(Mullis et al., 2009). The purpose of including this subgroup was for consistency with previous research that has examined the heterogeneous cognitive profiles associated with poor performance in math (e.g., Jordan et al., 2003; Swanson and Jerman, 2006).

The final version of the PIRLS reading assessment included texts that spanned many genres, including literary texts (e.g., short stories or episodes with illustrations), informational texts (e.g., biographies), and narratives and expositions (e.g., scientific, geographical, and procedural texts that included text boxes, photographs, maps, or diagrams; Mullis et al., 2009). Plausible values (i.e., estimates of student ability) were used to address issues of biased statistical inferencing and to allow the use of standard statistical tools to estimate population characteristics (Wu, 2005). Overall reliability of all reading comprehension items were estimated within the range of $\alpha = 0.86\text{--}0.91$.

No MD or MD/RD

A final group of students were above the 10th percentile on the Number subsection of TIMSS—regardless of their reading ability.

Language

This variable denotes language of the test, the classroom instruction, and the student's home language, Chinese or English.

Student Background Characteristics

Drawing on previous research, we selected gender (e.g., Nosek et al., 2009; Pieng et al., 2016) and maternal education (Bradley and Corwyn, 2002) as control variables in this study. We also controlled for country because education systems and associated resources (e.g., sequence of, or approach to skills taught within a country's program or resources in school organizations within cities, districts, etc., and required or adopted school curricula) vary by country and age to ensure any cross-linguistic differences could not be explained by differences in maturation between language groups.

In the current study, students responded to the questions “when were you born” and “are you a boy or a girl,” and caregivers answered questions about maternal education. In order to simplify the analysis, the nine categories of mother's education in the TIMSS/PIRLS home survey were collapsed into low, middle, and high. The 7% of students with missing mother's education data were identified as their own category and were included in the analysis.

Analysis Approach

We employed chi-square tests of independence to determine if there were differences in the samples by language group (Chinese vs. English). Then, to investigate the main effect of language on math achievement (and corroborate previous research) standardized values of GSM and DD were regressed on control variables for country, age, sex, and maternal education, and a dummy variable for English (i.e., 1 = English, 0 = Chinese). An additional set of regression models addressed the purpose of our study by considering a set of dummy variables for math ability and language by math ability interactions in the analysis. We also compared

ordinary regression models to hierarchical linear models (HLM) with likelihood ratio tests because students were nested in schools.

RESULTS

Results from initial chi-square tests showed that the Chinese- and English-speaking samples were roughly comparable in terms of student background characteristics. The only exception was that considerably more English-speaking students (26%) came from families in which the mother earned high degrees in education compared to their Chinese counterparts (12%; $p < 0.001$). As per the study design, the percentages of students with MD and MD/RD in both samples were consistent. The results from the descriptive statistics suggest that the language groups in each country performed consistently the same on GSM and DD in respect to whether or not they were above or below the grand mean. However, there was considerably more variability in the English scores across countries, than the Chinese scores, partly because there were simply more countries that took the test in English ($n = 8$) compared to Chinese ($n = 2$). **Table 1** provides descriptive statistics related to student demographics and **Table 2** provides the mean scores on the GSM and DD subtests by subgroup.

Based on the multilevel structure of TIMSS data (i.e., students nested within specific schools), likelihood-ratio tests were conducted, comparing ordinary regression to HLM models in order to investigate whether a random intercept for school was needed. Because all of the tests were significant, random intercepts for schools were included in all models. As a result, HLM models emerged as the best fitting to the data in all analyses, which we then presumed was the most appropriate analytic method to investigate cross-linguistic differences in math performance as a function of mathematical ability (Rabe-Hesketh and Skrondal, 2005). However, because the multilevel data structure was not the focus of this investigation, we do not discuss the multilevel aspects of our results further. Instead, we focus on interpreting the variables of interest in this study.

Table 3 provides the results from four models: Model (1) GSM was regressed on “English” (i.e., a dummy variable: English = 1, Chinese = 0) and the control variables, (2) GSM was regressed on English and the MD and MD/RD by English interactions and the control variables, (3) DD was regressed on “English” and the control variables, and (4) DD was regressed on English and the MD and MD/RD by English interactions and the control variables.

Hierarchical linear modeling (HLM) analyses revealed a significant main effect of language on DD ($p < 0.01$; $\beta = -0.34$) and a borderline significant relationship between language and GSM ($p < 0.06$; $\beta = -0.21$) such that students who learned math in English were on average performing below students who learned in Chinese, (e.g., Peak, 1996, 1997). While there were no notable differences between Chinese- and English-speaking students with MD/RD, there were significant language by mathematics ability interaction effects, while controlling

TABLE 1 | Participants by Language.

	Chinese n	Chinese %	English n	English %
Australia	–		3,012	19
Taiwan	4,087	55	–	–
Hong Kong	3,364	45	52	<1
Ireland	–	–	3,702	23
Malta	–	–	1,402	9
Northern Ireland	–	–	2,009	13
Qatar	–	–	637	4
Saudi Arabia	–	–	23	<1
Singapore	–	–	4,932	31
Male	3,929	53	7,753	49
Female	3,522	47	8,016	51
Low maternal education	1,932	26	2,396	15
Mid maternal education	4,073	55	8,115	51
High maternal education	926	12	4,112	26
Missing maternal education	520	7	1,146	7
MD/RD	330	4	734	5
MD	416	6	842	5
No MD or RD	6,705	90	14,193	90
Total	7,451	100	15,769	100

for country, gender, maternal education, and age. English-speaking students with only MD performed considerably below Chinese-speaking students with MD on DD ($p < 0.001$; $\beta = 0.15$) and GSM ($p < 0.06$; $\beta = 0.09$); while English-speaking students as a whole were, on average, 0.34 of a standard deviation below their Chinese-speaking counterparts on DD, there was an additional negative effect (-0.15 of a standard deviation) for English-speaking students with poor demonstrated math ability. Or, in other words, as students approached the tail end of the distribution, the gap between English and Chinese performance widened. This result is notable given that Singapore, in South East Asia, was the largest contributor to the English-speaking sample (31%).

Finally, results related to the control variables echoed findings from previous research in that students from families with relatively high maternal education outperformed lower maternal education students, and developmental maturity (age) was related to achievement such that older students had higher average scores than younger students. Gender was not related to achievement. Finally, the four top performing countries were Singapore and Saudi Arabia and Hong Kong and Taiwan. The lowest five countries were all English speaking (Malta, Qatar, Ireland, Northern Ireland, and Australia). Additionally, consistent with the descriptive statistics, even when accounting for the control variables, there were small differences in math performance in Hong Kong compared to Singapore (when students took the test in Chinese), and wide variability across the English students by country. However, even when taking into account the effects of country, age, and maternal education, and language, the additional joint effect of language and demonstrated math ability was consistently associated with math achievement.

TABLE 2 | Mean scores on geometric shapes and data display by language.

	Chinese GSM	Chinese DD	English GSM	English DD
Australia			–0.2	–0.43
Taiwan	0.14	0.43		
Hong Kong	0.49	0.39	0.62	0.34
Ireland	–	–	–0.46	–0.45
Malta	–	–	–0.72	–0.6
Northern Ireland	–	–	0.2	0.08
Qatar	–	–	–1.21	–1.03
Saudi Arabia	–	–	–2.37	–1.97
Singapore	–	–	0.3	0.26
Male	0.31	0.41	–0.13	–0.18
Female	0.28	0.41	–0.15	–0.21
Low maternal education	0.22	0.24	–0.6	–0.59
Mid maternal education	0.29	0.45	–0.13	–0.18
High maternal education	0.66	0.76	0.22	0.13
Missing maternal education	0.02	0.1	–0.56	–0.59
MD/RD	–1.41	–1.18	–2.1	–2.09
MD	–0.96	–0.77	–1.69	–1.73
No MD or RD	0.46	0.56	0.05	0
Grand mean	0.3	0.41	–0.14	–0.19

GSM range = -4.62 – 3.45 and DD range = -5.45 – 4.67 .

DISCUSSION

We aimed to investigate the potential impact of Chinese and English numerical language on fourth-grade mathematics learning, especially for students who were underperforming in math. Consistent with previous research, our findings suggested that, on average, Chinese-speaking students have stronger math performance compared with their English-speaking counterparts (Peak, 1996, 1997; Mullis et al., 2016). However, we also found preliminary evidence of an additional gap between Chinese and English math performance for students who were relatively proficient in reading but had the poorest math ability.

There are several limitations to this study. First, all findings are bound to the respective conceptual definitions and development of the PIRLS and TIMSS measures and procedures, which naturally constrains our approach for investigating explanatory variables (MD, MD/RD). Second, the TIMSS measures may lack the sensitivity needed to detect subtle differences between students with MD and those with co-occurring MD/RD. These weaknesses are balanced by the fact that large-scale datasets such as PIRLS and TIMSS provide the opportunity to investigate the relationship between math ability and language at a scale that is inaccessible to most individual researchers.

Third, as mentioned, it is not possible to account for the differences in instructional approaches and curricular sequences for math that may vary as a function of language and culture. This limitation is somewhat mitigated by the inclusion of control variables for country and random intercepts for schools, which controls for unobserved classroom-level variables (Rabe-Hesketh and Skrondal, 2005). Additionally, a significant

TABLE 3 | Fixed effects estimates and variance-covariance estimates for models of the predictors of fourth-grade mathematics achievement (standardized geometric shapes and data display) on the TIMSS 2011 assessment.

	Model 1 GSM	Model 2 GSM	Model 3 DD	Model 4 DD
Taiwan	0.17	0.07	0.54***	0.45***
Hong Kong	0.66***	0.46***	0.64***	0.44***
Ireland	−0.23***	−0.35***	−0.01	−0.12***
Malta	−0.35***	−0.38***	−0.00	−0.03
Northern Ireland	−0.95***	−0.72***	−0.55***	−0.33***
Qatar	−2.12***	−1.28***	−1.47***	−0.63***
Saudi Arabia	0.52***	0.28***	0.70***	0.46***
Singapore	0.42***	0.21***	0.54***	0.32***
Female	−0.01	−0.01	−0.01	−0.00
Age	0.09***	0.08***	0.10***	0.10***
Mid maternal education	0.32***	0.20***	0.31***	0.19***
High maternal education	0.71***	0.51***	0.66***	0.46***
Missing maternal education	−0.02	0.01	−0.05	−0.01
English	−0.22+	−0.21++	−0.35**	−0.34**
MD	−	−1.34***	−	−1.27***
MD/RD	−	−1.71***	−	−1.66***
MD × English	−	−0.09++	−	−0.15***
MD/RD × English	−	−0.01	−	−0.07
Intercept	−1.27***	−0.84***	−1.47***	−1.05***
ψ	−1.41***	−1.79***	−1.52***	−1.92***
Θ	−0.17***	−0.32***	−0.16***	−0.31***

Standard errors range: 0.01–0.19, ψ = between school variance and Θ = within school variance. Each coefficient can be understood as the comparison between each named group and its respective reference group (Australia, males, low maternal education, and students with no MD or MD/RD), when controlling for the other variables in the model. +p < 0.10, ++p < 0.06, *p < 0.05, **p < 0.01, ***p < 0.001.

weakness of our study is that only one country from the entire database administered the tests in both languages; as such, language may be confounded with country and/or culture for this particular investigation. However, notably, Singapore, in South East Asia, was the largest contributor to the English-speaking sample (31%), which supports the possibility that cross-language differences in our study are due to language differences rather than cultural differences.

Fourth, although both TIMSS and PIRLS made considerable efforts to make sure that the assessments were comparable across languages, it is quite possible that there were significant differences between the tests in the two languages (Flores, 2016), which is an especially important consideration given that the TIMSS math problems were given in a language context—most items included written directions and/or word problems. The fact that the tests were written in English and translated into Chinese could have considerable advantages/disadvantages for students, with the additional possibility that translation effects could uniquely influence students in the MD and MD/RD groups relative to the students without any MD. One additional point to consider, however, is that the fact that items originally constructed in English version would theoretically give students who took the assessment in English an advantage, which was

not the case based on our findings. As such, we believe that the likelihood of problematic differences in test versions to have a minimal impact on performance.

Lastly, this study focused on cross-linguistic differences in math performance; however, it is quite possible that because the ability to solve a math problem is necessarily dependent on reading—e.g., comprehension of the directions or words in a word problem—there may be differences between Chinese and English math performance that are due to differences in orthographies (Perfetti et al., 1992) instead of, or in addition to, how numbers are represented and therefore processed in each language. However, the influence of orthographic differences on math performance was beyond the scope of this study. Yet, since we included an assessment of student reading skills, we were able to distinguish between students who seemed to have difficulty in math due to poor reading skills versus students who had domain-specific difficulties in math.

Despite all described limitations, several tentative conclusions can be drawn from this study. Our results corroborate previous research showing notable cross-linguistic differences in fourth-grade mathematics achievement between Chinese- and English-speaking students (e.g., Peak, 1996). The significant MD by language interaction (coupled with the non-significant MD/RD by language interaction) seen consistently across both the Geometric Shapes and Data Display domains does raise the possibility of a continuing negative effect of learning math in the English for students with the poorest demonstrated levels of math ability. Surprisingly, the interaction between ability and language was unique to the MD group (and not the MD/RD group). One explanation is that students with MD/RD struggle with math mainly because of their poor reading skills (e.g., they have trouble reading directions or understanding word problems). Thus, they are not slowed down or confused by the relative irregularity of the English number system, but are limited by their weaknesses that are relatively specific to reading. In contrast, students with MD alone, who demonstrate that they are more proficient in reading, presumably struggle with basic math skills such as retrieving, holding, and acquiring number information during simple arithmetic (Geary, 1993). One logical conclusion, therefore, is that students who demonstrate weaknesses specific to math would be negatively (English) or positively (Chinese) affected by the degree of obviousness of their numerical language, which places differential demands on the learner. However, reading difficulties, which presumably influence math difficulties in the context of reading word problems in both languages, might not interact with the numerical characteristics of language in the same way as domain specific difficulties in math.

However, our finding that the MD/RD group did not demonstrate equal or even lower math performance than the MD group may be a contrast to previous research that has honed in on math impairment (e.g., Jordan et al., 2003; Swanson and Jerman, 2006) and suggests demonstrated weaknesses in math and reading are tied to the same underlying cognitive mechanisms (e.g., phonological processing, working memory) and students with both MD and RD tend to have even more difficulty in math (due to more significant impairment) than students with MD alone. However, in our study, in which we

focused on the tail end of distribution of math performance in Chinese and English (and not specific math impairment), the effect of relatively ambiguous numerical language on math performance seemed to be the most (negatively) pronounced for the English-speaking students whose learning differences are specific to domain of math (not reading).

The MD by language interaction may shed even more light on findings from other comparative studies on English- and Chinese-speaking students. For example, according to responses from the international assessment, *Test for Schools*, which is a part of the *Program for International Student Assessment* (PISA, OECD, 2013), even the most disadvantaged 15-year-old Chinese students in Shanghai are outperforming middle and higher socioeconomic students in the U.S. This disparity in academic performance had been generally described as the result of country-specific differences in the areas of teacher content knowledge, dedication, and support (Friedman, 2013; OECD, 2013). However, variation in language (i.e., the degree of transparency of number naming systems) may explain variation in math performance (Miller et al., 1995), and, according to findings from this present investigation, this effect of language on math performance may be conditional on math ability.

MD-Specific Interactions

In this study, we investigated the role of language on math performance for students of varying math abilities. This study contributes to these findings by examining the role of math ability in differing linguistic environments. Difficulty in the area of math is not entirely uncommon and has been associated with the difficulty to master number skills (e.g., representing the meaning of numbers; Geary et al., 1999; Landerl et al., 2004). Logically speaking, weaknesses in number representation may be exacerbated by number naming systems that less transparently correspond to numerical magnitudes. The results from this study provide cautious support for the hypothesis that cross-national differences (in both geometry and data analysis) in performance may be due in part to the obviousness (or lack thereof) of number naming systems that continues to be an obstacle for students with the poorest ability. This MD-specific interactions suggests that the students with the poorest ability, who might be on the tail end of the distribution in terms of their ability to represent numbers, counts, and manipulate mathematical information, may be uniquely challenged by languages that less transparently correspond to mathematical concepts (i.e., English). Thus, it may be informative for researchers and educators to look at particular subgroups of learners when considering cross-national and cross-linguistic differences in achievement, and that countries that lag behind Asian countries may consider specific changes in practice that target underperforming learners. For example, it may be particularly useful for students who are struggling in math in English to engage in on-going learning activities that strengthen knowledge of how (irregular) two-digit number names map on to numerical magnitudes according to the base-10 system (Zhang and Okamoto, 2017); and, teachers can be mindful of how early ease or difficulty with the acquisition of number names and their corresponding magnitudes in English, may continue to play a role in learning more advanced mathematical concepts in, such as in geometry or data analysis. For example, a solid

understanding of the underlying base-10 structure of decimals such as (0.90) may be the necessary foundation for learning probability and statistical inference. Likewise, awareness of the underlying morphological structure of English words, such as “bi,” “tri,” and “quad,” may be a prerequisite that dispels confusion around basic concepts and supports understanding of more complex concepts in geometry.

The Obviousness of Number Naming Systems

Less transparent number naming systems have been shown to inhibit math skills in for broad populations of children. The findings from this study show that that such opaque number systems may be specifically more cognitively demanding for student with poor math ability compared to systems that are more straightforward. Such variability in number words appears to be related to number representation, counting, and the ability to manipulate numerical information—which support higher-level math skills (such as geometry and data analysis). The word “rectangle,” for example, is the proper English representation of a long square shape, while in Chinese, the word for this shape is also a clear description; its Chinese counterpart, 长方形 is literally translated as “long square shape.”

Previous research that has suggested that cross-national disparities in achievement outcomes cannot be completely attributed to differences in educational systems (e.g., U.S. versus China); and, that instruction to address student weaknesses could focus on making the base-10 structure of number names more readily accessible to students (Miller et al., 1995). This study augments these findings by further specifying that efforts to support students should also target students with the poorest math ability. Future investigations on the effect of language on math performance should include the varying levels of math ability. Further explorations, perhaps more qualitative in nature, might be helpful in unpacking the observed differences in mathematical performance for Chinese speaking students in Taiwan and Hong Kong; country-level differences may have more to do with differences in educational standards and practices.

As educators, researchers and other scholars continue to investigate the differences in math performance across the world, the catalytic factors for varying levels of performance will undoubtedly be revealed. Demonstrating or using one's math knowledge is impossible without language and, as revealed in this study, specific difficulties in the domain of math may play a determining role in how much one's language becomes a hurdle (e.g., English) or springboard (e.g., Chinese) for demonstrating such knowledge. Understanding the potential roadblocks and supports for students as they continue to develop math knowledge and skills will ultimately benefit learning and instructional practice, regardless of how one counts out loud.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

- Arya, D. J., McClung, N. A., Katznelson, N., and Scott, L. (2015). Language ideologies and literacy achievement: six multilingual countries and two international assessments. *Int. J. Multilingual.* 13, 40–60. doi: 10.1080/14790718.2015.1021352
- Bolger, D. J., Perfetti, C. A., and Schneider, W. (2005). Cross-cultural effect on the brain revisited: universal structures plus writing system variation. *Hum. Brain Mapp.* 25, 92–104. doi: 10.1002/hbm.20124
- Bradley, R. H., and Corwyn, R. F. (2002). Socioeconomic status and child development. *Annu. Rev. Psychol.* 53, 371–399. doi: 10.1146/annurev.psych.53.100901.135233
- Caravolas, M. (2005). “The nature and causes of dyslexia in different languages,” in *The Science of Reading: A Handbook*, eds M. J. Snowling and C. Hulme (Oxford: Blackwell), 336–355.
- Flores, G. S. (2016). *Assessing English Language Learners: Theory and Practice*. New York, NY: Routledge.
- Friedman, T. L. (2013). *My Little (Global) School*. Available online at: <https://www.nytimes.com/2013/04/03/opinion/friedman-my-little-global-school.html>
- Frost, R. (2005). “Orthographic systems and skilled word recognition processes in reading,” in *The Science of Reading: A Handbook*, eds M. J. Snowling and C. Hulme (Oxford: Blackwell), 272–295.
- Fuson, K. C., and Kwon, Y. (1992). “Effects of the system of number words and other cultural tools on children’s addition and subtraction,” in *Pathways to Number: Children’s Developing Numerical Abilities*, eds J. Bideaud and C. Maljac (Hillsdale, NJ: Erlbaum), 127–149.
- Geary, D. C. (1993). Mathematical disabilities: cognitive, neuropsychological and genetic components. *Psychol. Bull.* 114, 345–362. doi: 10.1037/0033-2909.114.2.345
- Geary, D. C., Bow-Thomas, C. C., and Yao, Y. (1992). Counting knowledge and skill in cognitive addition: a comparison of normal and mathematically disabled children. *J. Exp. Child Psychol.* 54, 372–391. doi: 10.1016/0022-0965(92)90026-3
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., and DeSoto, M. C. (2004). Strategy choices in simple and complex addition: contributions of working memory and counting knowledge for children with math disability. *J. Exp. Child Psychol.* 80, 121–151. doi: 10.1016/j.jecp.2004.03.002
- Geary, D. C., Hoard, M. K., and Hamson, C. O. (1999). Numerical and arithmetical cognition: patterns of functions and deficits in children at risk for a mathematical disability. *J. Exp. Child Psychol.* 74, 213–239. doi: 10.1006/jecp.1999.2515
- Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., and Nuerk, H. C. (2014). Language affects symbolic arithmetic in children: The case of number word inversion. *J. Exp. Child Psychol.* 119, 17–25. doi: 10.1016/j.jecp.2013.10.001
- Goldman, S. R., Pellegrino, J. W., and Mertz, D. L. (1988). Extended practice of basic addition facts: strategy changes in learning disabled students. *Cogn. Instr.* 5, 223–265. doi: 10.1207/s1532690xci0503_2
- Hanley, R., Masterson, J., Spencer, L., and Evans, D. (2004). How long do the advantages of learning to read a transparent orthography last? An investigation of the reading skills and reading impairment of Welsh children at 10 years of age. *Q. J. Exp. Psychol.* 57, 1393–1410. doi: 10.1080/02724980343000819
- Helmreich, I., Zuber, J., Pixner, S., Kaufmann, L., Nuerk, H.-C., and Moeller, K. (2011). Language effects on children’s nonverbal number line estimations. *J. Cross Cult. Psychol.* 42, 598–613. doi: 10.1177/0022022111406026
- Hiebert, J., Gallimore, R., Garnier, H., Givvin, K., Hollingsworth, H., Jacobs, J., et al. (2003). *Teaching Mathematics in Seven Countries: Results from the TIMSS 1999 Video Study*. NCES 2003-013. Washington, DC: National Center for Education Statistics.
- Hiebert, J., Stigler, J. W., Jacobs, J. K., Givvin, K. B., Garnier, H., Smith, M., et al. (2005). Mathematics teaching in the United States today (and tomorrow): results from the TIMSS 1999 video study. *Educ. Eval. Policy Anal.* 27, 111–132. doi: 10.3102/01623737027002111
- Hurford, J. R. (1975). *The Linguistic Theory of Numerals*. Cambridge: Cambridge University Press.
- Hurford, J. R. (1987). *Language and Number*. Cambridge: Cambridge University Press.
- Husen, T. (1967). *International Study of Achievement in Mathematics*. New York, NY: Wiley.
- Imbo, I., Bulcke, C. V., De Brauwier, J., and Fias, W. (2014). Sixty-four or four-and-sixty? The influence of language and working memory on children’s number transcoding. *Front. Psychol.* 5:313. doi: 10.3389/fpsyg.2014.00313
- Jordan, N. C., Hanich, L. B., and Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with co-morbid mathematics and reading difficulties. *Child Dev.* 74, 834–850. doi: 10.1111/1467-8624.00571
- Jordan, N., and Montani, T. (1997). Cognitive arithmetic and problem solving: a comparison of children with specific and general mathematics difficulties. *J. Learn. Disabil.* 30, 624–634. doi: 10.1177/002221949703000606
- Landerl, K., Bevan, A., and Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8–9 year old students. *Cognition* 93, 99–125. doi: 10.1016/j.cognition.2003.11.004
- Martin, M. O., and Mullis, I. V. S. (eds.). (2012). *Methods and procedures in TIMSS and PIRLS 2011*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Miller, K. F., Smith, C. M., and Zhang, H. (2004). *Language and Number: A Longitudinal Study of Learning to Count in Chinese and English*. Unpublished manuscript, University of Illinois at Urbana-Champaign.
- Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: The role of number-naming systems. *Psychol. Sci.* 6, 56–60.
- Miller, K. F., and Stigler, J. (1987). Counting in Chinese: cultural variation in basic cognitive skill. *Cogn. Dev.* 2, 279–305. doi: 10.1016/S0885-2014(87)90091-8
- Miller, K., Kelly, M., and Zhou, X. (2005). “Learning mathematics in China and the United States,” in *Handbook of Mathematical Cognition*, ed J. Campbell (New York, NY: Psychology Press), 163–178.
- Miura, I. T. (1987). Mathematics achievement as a function of language. *J. Educ. Psychol.* 79, 79–82. doi: 10.1037/0022-0663.79.1.79
- Miura, I. T., Kim, C. C., Chang, C., and Okamoto, Y. (1988). Effects of language characteristics on children’s cognitive representation of number: cross-national comparisons. *Child Dev.* 59, 1445–1450. doi: 10.2307/1130659
- Miura, I. T., and Okamoto, Y. (1989). Comparisons of U.S. and Japanese first graders’ cognitive representation of number and understanding of place value. *J. Educ. Psychol.* 81, 109–113. doi: 10.1037/0022-0663.81.1.109
- Moeller, K., Shaki, S., Göbel, S. M., and Nuerk, H. C. (2015). Language influences number processing—a quadrilingual study. *Cognition* 136, 150–155. doi: 10.1016/j.cognition.2014.11.003
- Mullis, I. V. S., Martin, M. O., Foy, P., and Hooper, M. (2016). *TIMSS Advanced 2015 International Results in Advanced Mathematics and Physics*. Boston College, TIMSS & PIRLS International Study Center. Available online at: <http://timssandpirls.bc.edu/timss2015/international-results/advanced/>
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., Trong, K. L., and Sainsbury, M. (2009). *PIRLS 2011 Assessment Framework*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., et al. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10593–10597. doi: 10.1073/pnas.0809921106
- OECD (2010). *PISA 2009 Results: Executive Summary*.
- OECD (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*. OECD Publishing. doi: 10.1787/9789264190511-en
- Okamoto, Y. (2017). “Effects of mathematics language on children’s achievement and central conceptual knowledge,” in *Language and Culture in Mathematical Cognition*, ed D. Berch (San Diego, CA: Elsevier).
- Passolunghi, M. C., and Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *J. Exp. Child Psychol.* 88, 348–367. doi: 10.1016/j.jecp.2004.04.002
- Peak, L. (1996). *Pursuing Excellence: A Study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context: Initial Findings From The Third International Mathematics and Science Study (1995)*. Institute of Educational Sciences (IES). Available online at: <http://nces.ed.gov/pubs97/timss/97198.asp>
- Peak, L. (1997). *Pursuing Excellence. A Study of U.S. Eight-Grade Mathematics and Science Achievement in an International Context*. Washington, DC: U.S. Government Printing Office.

- Perfetti, C. A., Liu, Y., and Tan, L. H. (2005). The lexical constituency model: some implications of research on Chinese for general theories of reading. *Psychol. Rev.* 112, 43–59 doi: 10.1037/0033-295X.112.1.43
- Perfetti, C. A., Zhang, S., and Berent, I. (1992). “Reading in English and Chinese: evidence for a “universal” phonological principle,” in *Advances in Psychology: Orthography, Phonology, Morphology, and Meaning*, eds R. Frost and L. Katz (Oxford: NorthHolland), 227–248.
- Pieng, P., Okamoto, Y., and La Joy, J. (2016). “Exploring gender differences and developmental changes in Japanese adolescent and young adult self-concept,” in *Handbook of Comparative and International Studies in Education*, ed D. Sharps, A volume in the series: *International Perspectives on Educational Policy, Research and Practice*, ed G. E. Fischman (Charlotte, NC: Information Age Publishing), 495–514.
- Rabe-Hesketh, S., and Skrondal, A. (2005). *Multilevel and Longitudinal Modeling Using Stata*, 3rd Edn. College Station, TX: Stata Press.
- Rousselle, L., and Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: a comparison of symbolic vs. non-symbolic number magnitude. *Cognition* 102, 361–395. doi: 10.1016/j.cognition.2006.01.005
- Seymour, P., Aro, M., and Erskine, J. (2003). Foundation literacy acquisition in European orthographies. *Br. J. Psychol.* 94, 143–174. doi: 10.1348/000712603321661859
- Siok, W. T., Perfetti, C. A., Jin, Z., and Tan, L. H. (2004). Biological abnormality of impaired reading is constrained by culture. *Nature* 431, 71–76. doi: 10.1038/nature02865
- Stevenson, H. W., Lee, S., and Rohlen, T. (1998). An examination of American student achievement from an international perspective. *Brookings Pap. Educ. Policy* 1, 7–52.
- Stevenson, H. W., Lee, S. Y., and Stigler, J. W. (1986). Mathematics achievement of Chinese, Japanese, and American children. *Science* 231, 693–699. doi: 10.1126/science.3945803
- Stevenson, H. W., and Stigler, J. W. (1992). *The Learning Gap*. New York, NY: Summit Books.
- Stigler, J. W., Gallimore, R., and Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: examples and lessons from the TIMSS video studies. *Educ. Psychol.* 35, 87–100. doi: 10.1207/S15326985EP3502_3
- Stigler, J. W., Lee, S. Y., Lucker, G. W., and Stevenson, H. W. (1982). Curriculum and achievement in mathematics: a study of elementary school children in Japan, Taiwan and the United States. *J. Educ. Psychol.* 74, 315–322. doi: 10.1037/0022-0663.74.3.315
- Swanson, H. L., and Jerman, O. (2006). Math disabilities: a selective meta-analysis of the literature. *Rev. Educ. Res.* 76, 249–274. doi: 10.3102/00346543076002249
- Tan, L. H., Laird, A., Li, K., and Fox, P. T. (2005). Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: a meta-analysis. *Hum. Brain Mapp.* 25, 83–91. doi: 10.1002/hbm.20134
- TIMSS (2011). *Assessment*. Copyright © 2012 International Association for the Evaluation of Educational Achievement (IEA). Chestnut Hill, MA; Amsterdam: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College; International Association for the Evaluation of Educational Achievement (IEA), IEA Secretariat.
- Travers, K. J., Crosswhite, F. J., Dossey, J. A., Swafford, J. O., McKnight, C. C., and Cooney, T. J. (1987). *Second International Mathematics Study Summary Report for the United States*. Champaign, IL: Stipes.
- Vellutino, F. R. (1979). *Dyslexia: Theory and Research*. Cambridge, MA: MIT press.
- Wu, M. (2005). The role of plausible values in large-scale surveys. *Stud. Educ. Eval.* 31, 114–128. doi: 10.1016/j.stueduc.2005.05.005
- Zhang, Y., and Okamoto, Y. (2017). Encoding “10ness improves first graders” estimation of numerical magnitudes. *J. Numerical Cogn.* 2, 202–219. doi: 10.5964/jnc.v2i3.69

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

The reviewer BDF and handling Editor declared their shared affiliation.

Copyright © 2018 McClung and Arya. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Mathematical (Dis)abilities Within the Opportunity-Propensity Model: The Choice of Math Test Matters

Elke Baten^{1*} and Annemie Desoete^{1,2}

¹ Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium, ² Department of Speech and Language Pathology, University College Arteveldehogeschool, Ghent, Belgium

This study examined individual differences in mathematics learning by combining antecedent (A), opportunity (O), and propensity (P) indicators within the Opportunity-Propensity Model. Although there is already some evidence for this model based on secondary datasets, there currently is no primary data available that simultaneously takes into account A, O, and P factors in children with and without Mathematical Learning Disabilities (MLD). Therefore, the mathematical abilities of 114 school-aged children (grade 3 till 6) with and without MLD were analyzed and combined with information retrieved from standardized tests and questionnaires. Results indicated significant differences in personality, motivation, temperament, subjective well-being, self-esteem, self-perceived competence, and parental aspirations when comparing children with and without MLD. In addition, A, O, and P factors were found to underlie mathematical abilities and disabilities. For the A factors, parental aspirations explained about half of the variance in fact retrieval speed in children without MLD, and SES was especially involved in the prediction of procedural accuracy in general. Teachers' experience contributed as O factor and explained about 6% of the variance in mathematical abilities. P indicators explained between 52 and 69% of the variance, with especially intelligence as overall significant predictor. Indirect effects pointed towards the interrelatedness of the predictors and the value of including A, O, and P indicators in a comprehensive model. The role parental aspirations played in fact retrieval speed was partially mediated through the self-perceived competence of the children, whereas the effect of SES on procedural accuracy was partially mediated through intelligence in children of both groups and through working memory capacity in children with MLD. Moreover, in line with the componential structure of mathematics, our findings were dependent on the math task used. Different A, O, and P indicators seemed to be important for fact retrieval speed compared to procedural accuracy. Also, mathematical development type (MLD or typical development) mattered since some A, O, and P factors were predictive for MLD only and the other way around. Practical implications of these findings and recommendations for future research on MLD and on individual differences in mathematical abilities are provided.

Keywords: Opportunity-Propensity Model, Mathematical Learning Disabilities, temperament, personality, motivation, subjective well-being, self-esteem, self-perceived competence

OPEN ACCESS

Edited by:

Yvette Renee Harris,
Miami University, United States

Reviewed by:

James P. Byrnes,
Temple University, United States
Jessica S. Horst,
University of Sussex, United Kingdom

*Correspondence:

Elke Baten
Elke.Baten@UGent.be

Specialty section:

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

Received: 11 August 2017

Accepted: 17 April 2018

Published: 08 May 2018

Citation:

Baten E and Desoete A (2018)
Mathematical (Dis)abilities Within
the Opportunity-Propensity Model:
The Choice of Math Test Matters.
Front. Psychol. 9:667.
doi: 10.3389/fpsyg.2018.00667

INTRODUCTION

Mathematical competence relies on several interrelated mechanisms and skills (Siemann and Petermann, 2018). Procedural skills are required to understand principles and solve calculations in a number problem (e.g., $48 + 6 = \dots$) or in a word problem (e.g., 6 more than 48 is ...) format. Additionally, mathematical competence relies on the capacity to remember and retrieve arithmetic facts (e.g., $16 : 4 = \dots$) with ease. Therefore, mathematics is considered to be componential in nature (Dowker, 2015). Research shows a lot of individual variation in school-taught mathematical abilities from the first year of primary school onwards (Mooij and Driessen, 2008; Clements and Sarama, 2011; Schuchart et al., 2015). To provide insight into the nature of these differences, some studies focused on predictors for mathematical outcomes, whereas others have compared children with and without Mathematical Learning Disabilities (MLD). MLD is a neurodevelopmental disorder characterized by mathematic skills substantially lower than expected with regard to the individual's chronological age and by persisting math problems despite interventions that target those difficulties (Bryant et al., 2015; Pieters et al., 2015; Baten et al., 2017). Worldwide, the prevalence of MLD is estimated between 5 and 7% (Shalev, 2007; Shin and Bryant, 2015). In addition, some authors propose that MLD is a heterogeneous disability with a procedural and a semantic memory subtype (Henik et al., 2015). The procedural subtype is characterized by a delay in the acquisition of procedural calculation procedures. In contrast, the semantic memory subtype is marked by a lack of fact retrieval fluency (Pieters et al., 2015).

Previous research focused on domain-specific cognitive predictors of mathematics, such as symbolic numerical processing (Vanbinst et al., 2015) and seriation and classification (Stock et al., 2010) in pre-school. In addition, studies demonstrated the relationship between domain-general cognitive abilities such as intelligence (Desoete, 2008; Dix and van der Meer, 2015) and working memory (De Weerd et al., 2013) on the one hand, and mathematical abilities on the other hand. Moreover, socioeconomic status (SES; Jordan and Levine, 2009; Aunio and Niemivirta, 2010) and parental academic aspirations (Murayama et al., 2016) were studied as contextual predictors. Finally, some researchers focused on non-cognitive predictors such as personality (e.g., Poropat, 2009) and motivation (Ryan and Deci, 2000; Froiland and Worrell, 2016).

However, by focusing on single predictors, the importance and unique explained variance of these predictors could be overestimated. Surprisingly few studies have been conducted to explore the combined effect of predictors. This study addresses this gap by investigating multiple predictors at the same time, within a comprehensive model to get a more holistic insight on math development. In what follows, we describe the model that will be used.

Byrnes and Miller (2007) developed the Opportunity-Propensity (O-P) framework, aiming to differentiate between opportunity (O) and propensity (P) factors in an effort to explain variance and individual differences in development. P factors are variables that make people able (e.g., intelligence) and/or

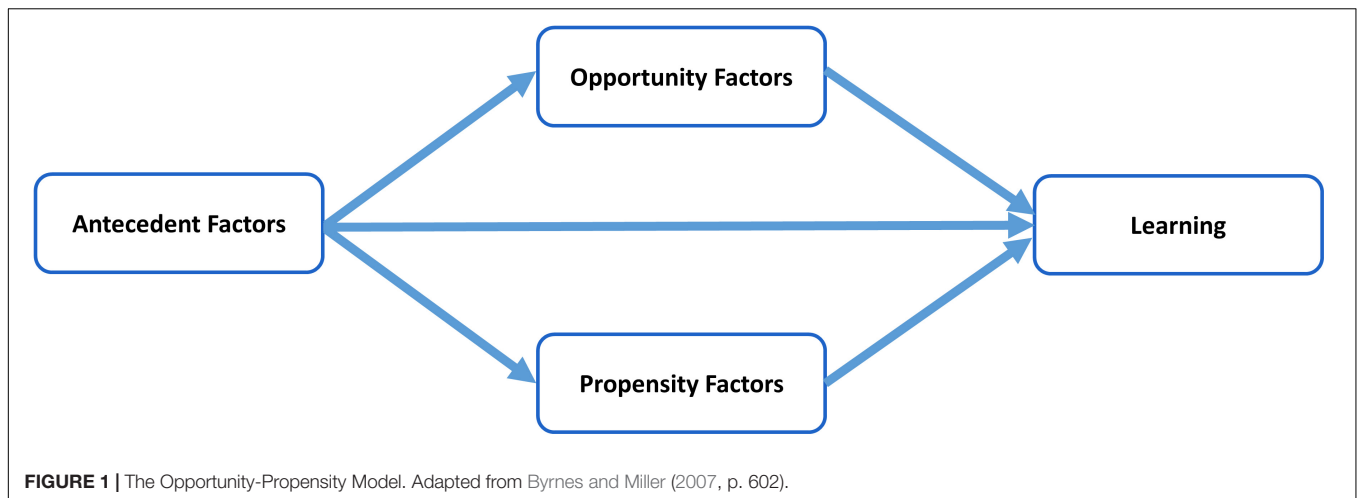
willing (e.g., motivation) to learn. O factors include contexts and variables that expose children to learning content (e.g., home environment, classroom instruction). Antecedent (A) or distal variables, for example SES, are present early in a child's life and explain why some people are exposed to richer O contexts and have stronger P's for learning than others (Byrnes and Miller, 2007, 2016; Wang and Byrnes, 2013; Ceulemans et al., 2017). A visual representation of the model can be found in **Figure 1**.

The O-P Model has been tested by using secondary datasets. However, these studies are still scarce, since there are only three studies about this. In the first longitudinal study, researchers explained about 80% of variance through A, O, and P factors in secondary school children in the United States who were followed from 8th up until 10th grade. Path analysis confirmed the causality between A factors on the one hand and O and P factors on the other hand, as well as causality between the latter two and math achievement. Although the effect of A factors was mediated through O and P factors, A factors had a direct but small effect on math results (Byrnes and Miller, 2007). A second longitudinal study with data from kindergarten up until primary school revealed additional evidence for the O-P Model with P factors as the strongest predictors (Byrnes and Wasik, 2009). Finally, Wang et al. (2013) found evidence for this model in lower-income pre-kindergarten children. Using Structural Equation Modeling (SEM), it was confirmed that the latent A factors predicted both the latent O and P factors and the latent O factor predicted early math skills. The predictive value of the latent P factor was not confirmed. However, significant predictions for early math skills could be made based on intelligence and self-regulation as P factors.

Because the current study intends to combine variables in an O-P Model and because there are to the best of our knowledge only three studies combining different predictors (see previous paragraph), the results of research examining these variables separately will be summarized here. Furthermore, all variables will be categorized as A, O, or P variables.

Studies including A indicators, revealed the role of SES in math development, especially in low-income families (Wang et al., 2013). Moreover, parental stimulation has also been related to mathematical achievement, although it remains unclear whether this relation was direct or mediated through intelligence or the availability of certain resources such as books, computers, etc. (Blevins-Knabe et al., 2007; Kleemans et al., 2012; Niklas et al., 2016). In addition, lower birth weight was related to lower levels of math performance at school-age level, with especially strong effects for extremely low birth weight (<1500 g; De Rodrigues et al., 2006; Chatterji et al., 2014). Finally, children who are born first seem to perform better in academic contexts. This has been explained by the dilution hypothesis in which the first born child takes advantage of more parental resources (at least for the time the child is an only child), compared to later born children who have to share these resources (Hotz and Pantano, 2015).

Studies including P indicators demonstrated that motivation, personality, temperament, intelligence, and working-memory capacity as well as well-being variables, predicted mathematics. In a meta-analysis on 18 studies, Taylor et al. (2014) highlighted a positive relationship between autonomous motivation (where



the force to fulfill a task is internal, e.g., passion) and general school achievement, in addition to a negative relationship between controlled motivation (where the force to fulfill a task is external, e.g., rewards-related) and academic achievement. According to research on the Big Five Personality Theory (Costa and McCrae, 1992), conscientiousness and openness are the personality traits most strongly associated with better academic performances, even when controlling for intelligence (Poropat, 2009; Zhang and Ziegler, 2016). Furthermore, math performance correlated positively with emotional stability (Zhang and Ziegler, 2016). Temperament, which is considered as the biological base of personality and described by the Reward Sensitivity Theory (Gray, 1981) as mechanisms guiding human behavior in terms of reactivity and self-regulation, can be seen as a P factor. More specifically, the unique constellation of one's temperament could make people willing and able to learn (Van Beek et al., 2013). Research on 565 Dutch University students revealed that pursuing rewards or positive consequences (higher Behavioral Activation System – BAS) was associated with higher study engagement and better academic performances. A temperament characterized by trying to avoid punishment or negative consequences (higher Behavioral Inhibition System – BIS) was related to more overcommitment and lower academic performances through exhaustion (Van Beek et al., 2013). Studies on intelligence and working-memory showed positive correlations with mathematical abilities (Roth et al., 2015; Peng and Fuchs, 2016). Moreover, well-being can be considered a P factor, since it makes people willing and able to learn. Positive and bidirectional relations between subjective well-being (SWB) and academic performance were found. For instance, Quinn and Duckworth (2007) revealed in 257 fifth grade students that higher levels of SWB (indicated by high levels of life satisfaction as cognitive component; and more positive emotions than negative emotions as affective component) were related to better academic performance and vice versa. This relationship was significant even when controlling for intelligence. Furthermore, higher perceptions of own academic competence were predictive of better academic achievement

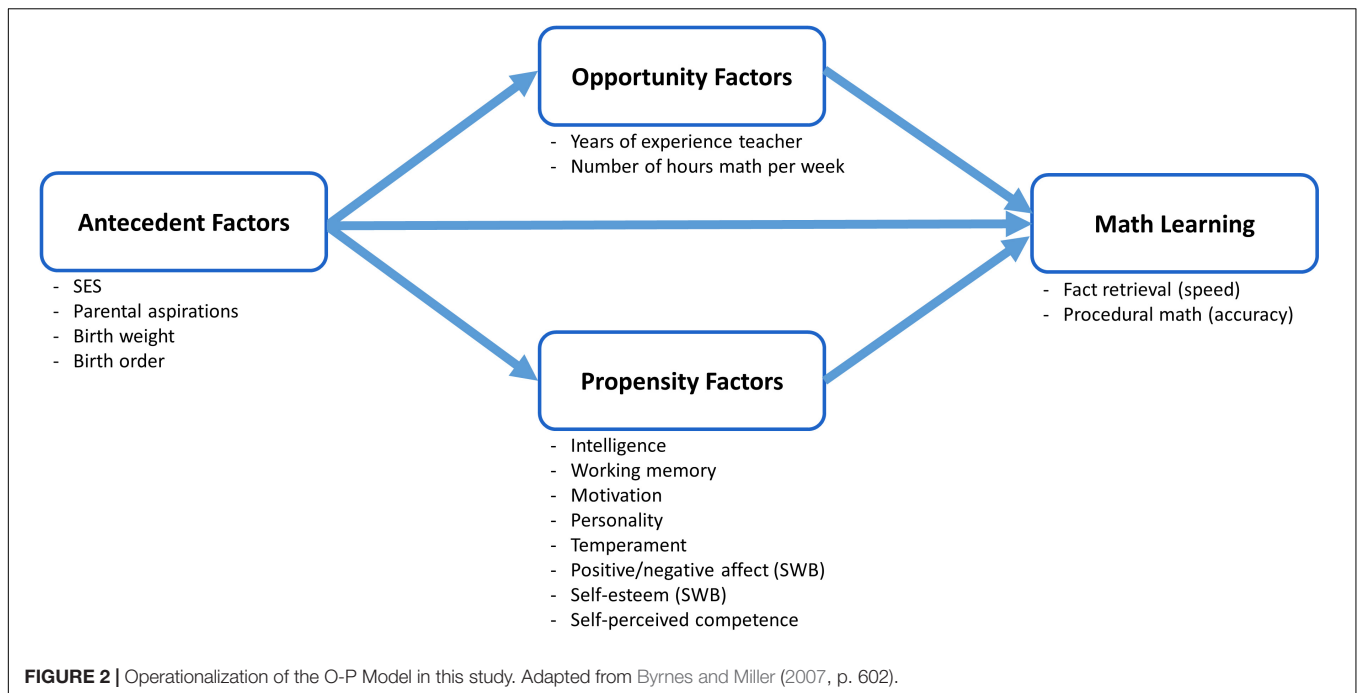
and the other way around (Arefi et al., 2014) which confirmed the reciprocal-effects model between academic self-concept and academic achievement (Guay et al., 2003; Seaton et al., 2015).

As to the O factors, teaching methods (Savelsbergh et al., 2016), instructional time (Cattaneo et al., 2016), teacher education level, and teachers' years of experience (Zhang, 2008) were found to be responsible for more O's to learn. The impact of the O factors depended on the specific support factors (Byrnes and Wasik, 2009; Cowan, 2015).

The Current Study

This study aimed to add some nuance to the literature on individual differences in mathematics learning by combining A, O, and P indicators within the O-P Model. Although there is already some evidence for this model (Byrnes and Miller, 2016, 2007; Byrnes and Wasik, 2009; Wang and Byrnes, 2013) from secondary datasets, there is little research from primary data simultaneously tapping the A's, O's, and P's empirically in children with and without MLD. Therefore, this study had the objective to extend the literature on the O-P Model in several ways. First, a variety of non-cognitive variables that had not yet been investigated in the context of this theory (e.g., temperament, personality, and self-perceived competence) were included. Second, the current study investigated specificity and examined differences between children with and without MLD on A's and P's and explored if there were different relationships with outcome depending on group (MLD or control). As such, this study contributes to theory-building about mathematical learning since it investigates whether the same learning models can be applied for children with and without clinical diagnosis. Finally, this study expands previous findings by taking the componential nature of mathematics into account by separately examining the prediction for procedural calculation and fact retrieval skills among children (Cohen Kadosh and Dowker, 2015; Pieters et al., 2015).

The operationalization of the O-P Model in the current study is described in **Figure 2**. Four major hypotheses were examined:



- (1) There will be differences in A and P indicators between children with and without MLD. We expect children with and without MLD to differ on these specific variables in a way that the negative predictors (Hypotheses 2 and 3) for math performance will be higher in children with MLD and vice versa.
- (2) A selection of A, O, and P indicators will predict fact retrieval speed.
- (3) A selection of A, O, and P indicators will predict procedural accuracy.
- (4) The predictive value of some A variables will be mediated through some O and P variables.

For Hypotheses 2 and 3, based on the literature described above, it is expected that children will have better mathematical abilities when they are raised with more O's (Zhang, 2008; Byrnes and Wasik, 2009), higher levels of SES (Wang et al., 2013), higher parental aspirations (Blevins-Knabe et al., 2007; Kleemans et al., 2012; Niklas et al., 2016), were born with higher birth weight (De Rodrigues et al., 2006; Chatterji et al., 2014) and have a higher place in the birth order (Hotz and Pantano, 2015). Furthermore, higher levels of autonomous motivation (Taylor et al., 2014), conscientiousness and openness (personality; Poropat, 2009; Zhang and Ziegler, 2016), BAS (temperament; Van Beek et al., 2013), positive affect and self-esteem (SWB; Quinn and Duckworth, 2007), self-perceived competence (Arefi et al., 2014), and intelligence and working memory (Roth et al., 2015; Peng and Fuchs, 2016) are expected to positively predict mathematical abilities. On the contrary, variables such as less O's, lower levels of SES, lower parental aspirations, lower birth weight, a lower place in the birth order, and lower levels of emotional stability are expected to be associated with lower levels of mathematical performance. Also higher levels of controlled

motivation, BIS (temperament) and negative affect (SWB) are supposed to result in lower math performances. Since this has never been explicitly investigated, no specific hypotheses are made for the different components (procedural calculation and fact retrieval) of mathematics.

MATERIALS AND METHODS

Sample

This study was conducted on 114 children (79 females) from 3rd up until 6th grade in Flanders. There were 61 children in the MLD group and 53 children were recruited from the same classrooms to be a part of the control group. This was done to maximize the possibility that the O factors at school level (school learning environment) were the same in both groups. When recruiting someone from the same class was not possible (in 22.9% of the sample), a matched participant was selected based on age, grade, and gender.

All children in the MLD group met the criteria for MLD, and performed below average (substantially and quantifiably, below the 16th percentile), while performance was resistant to instruction (Ghesquière, 2014). Comorbidity with reading disabilities, Attention Deficit Hyperactivity Disorder (ADHD) and Developmental Coordination Disorder (DCD) was allowed, because of the high comorbidity rates with MLD (Scheiris and Desoete, 2008; Pieters et al., 2009, 2015; Kucian and von Aster, 2015). The mean intelligence (see section "Material") was significantly lower in the MLD group ($M = 91.119$; $SD = 1.508$) compared to the control group ($M = 103.359$; $SD = 1.614$), $F(1,115) = 30.725$, $p < 0.001$. For SES, as measured by the Hollingshead Index (see section "Material"), there were

no significant differences between groups, $F(1,115) = 0.320$, $p = 0.573$. Mean SES in the MLD group was 42.913 ($SD = 10.512$), and for the control group, the mean SES was 44.021 ($SD = 10.577$).

Children were recruited by spreading flyers, through social media, schools, psychologists, and language and speech therapists in Flanders. Children's parents agreed for the research by signing an informed consent. This research was approved by the Ethical Committee of the Faculty of Psychology and Educational Sciences of Ghent University.

Procedure

After parents agreed to the participation of their children, two appointments for the actual research were made. Each session lasted about 90 min while tests and questionnaires were administered individually for each child. For some participants, recent test data (max. 1.5 years) for intelligence and mathematics was already available from, for example, their psychologist. In that case, the available data (measured with the same tests as in this study) was used to prevent test-retest effects. Testing happened in a location chosen by the parents. Most often, this was the school or at home. The researcher gave standardized instructions and was available to answer questions. The first session started with the fact retrieval test. After that, intelligence and working memory were measured, followed by completing the questionnaires. The procedural accuracy math test was completed in the second session together with the remaining questionnaires. The specific order in which the questionnaires were filled out, could not be fully standardized, due to lot of individual differences between children regarding the duration of the standardized tests and their alertness during research. Therefore, the order was adapted to keep the child motivated to take part in the research by, for example, alternating longer with shorter questionnaires.

The questionnaires for the parents and the teacher were given to the parents during the first session, and handed back to the researchers after the research had finished.

Material

Antecedent (A) and O factors were measured through questionnaires. More specifically, for the O factors, teachers were asked how many *years of experience* they had with teaching mathematics and how many hours of mathematical instructions the children received per week (*teaching hours*).

To measure A factors, parents were asked about their aspirations regarding the mathematical abilities of their children. They had to reflect on the score they wanted their child to have at the end of the current school year (in percentage). Additionally, information on *birth order* and *birth weight* of the child was collected. The SES of the family of the child was calculated using the Hollingshead index, combining the educational level and the current job of both parents into one score. The higher this score, the higher the SES of the family (Hollingshead, 1975, Unpublished). With regards to the P factors the following instruments were used.

Intelligence

It was measured using an abridged Dutch version of the Wechsler Intelligence Scale for Children-III (WISC-III-NL; Kort et al., 2005). The total intelligence quotient or IQ ($M = 100$; $SD = 15$) was obtained by combining the separate scores on the following subtests: Vocabulary, Similarities, Picture Concepts, and Block Design. The reliability of this short form was 0.92 and the distribution of total IQ-scores calculated with the short form did not significantly differ from the distribution of scores on the full intelligence test (Grégoire, 2000). Cronbach's α of the total IQ in the current sample was 0.795.

Working Memory

It was assessed with the Working Memory Index of the Dutch version of the Clinical Evaluation of Language Fundamentals-4 (CELF-IV-NL; Kort et al., 2008). By combining the subtests of Forward and Backward Number Repetition and the subtest of Familiar Sequences, a score for working memory was calculated. Cronbach's α was 0.786 for this sample.

Motivation for Mathematics

It was measured with the Dutch version of the Academic Self-Regulation Scale (Vansteenkiste et al., 2009) which consists of 24 questions which allow the calculation of the level of autonomous and controlled academic motivation. As suggested by the authors, the introduction for the questions was changed from "I am motivated to study because..." to "I am motivated to study mathematics because..." in order to measure motivation with regards to mathematics specifically. The child had to respond on a 5-point Likert scale to statements such as "because I find this an important goal in my life" as an index of autonomous motivation and "because other people (e.g., parents, friends, teachers) oblige me to do so" to measure controlled motivation. The score for each scale was calculated by averaging the score on the items belonging to that scale. Cronbach's α for this sample was 0.849 for autonomous and 0.727 for controlled motivation.

Personality

It was assessed by the Hierarchical Personality Inventory for Children (HiPIC; Mervielde and de Fruyt, 2009), filled out by the parents. This questionnaire was based on the Big Five Personality Theory (Costa and McCrae, 1992) and consisted of 144 items to measure the five personality traits: openness, conscientiousness, extraversion, agreeableness, and emotional stability (versus neuroticism). For each item, the parent had to indicate on a 5-point Likert scale how well that item applied to their child (e.g., "my child likes to learn new things"). The score for each personality trait was calculated using an algorithm in which some items were recoded inversely. The internal consistency of this questionnaire was good ($\alpha = 0.80-0.92$) with a test-retest reliability of $\alpha = 0.72-0.83$ (Egberink et al., 2010). Cronbach's α for this sample was 0.868 for openness, 0.920 for conscientiousness, 0.642 for extraversion, 0.686 for agreeableness, and 0.905 for emotional stability.

Temperament

It was estimated with the Behavioral Inhibition (BIS) and Behavioral Activation (BAS) Questionnaire (Carver and White,

1994; translated by Franken et al., 2005). The children were asked to rate 24 items on a 4-point Likert scale. The score for BIS was calculated by averaging the score on seven items, for example, “I worry about making mistakes.” Two out of seven items were recoded reversely. For BAS, the score was calculated by averaging the score on 13 items, for example, “When I want something, I usually go all-out to get it.” The internal consistency of the scales have proven to be acceptable with BIS: $\alpha = 0.82$ and BAS: $\alpha = 0.73$ (Smits and Boeck, 2006). Cronbach’s α for this sample was 0.625 for BIS and 0.752 for BAS.

Subjective Well-Being

It was determined through the Dutch version of the Positive and Negative Affect Schedule (PANAS; Watson et al., 1988; translated by Engelen et al., 2006). Children indicated on a 5-point Likert scale how many negative (e.g., guilt and sadness) and positive (e.g., success and interest) emotions they experienced on a regular school day. Scores were calculated for the level of positive affect and the level of negative affect by averaging the score on 10 items. Cronbach’s α for this sample was 0.738 for positive affect and 0.708 for negative affect.

Self-Esteem

It was evaluated through the Dutch version of the Rosenberg self-esteem scale (Franck et al., 2008). Children had to judge 10 statements on a 4-point Likert scale. Examples of questions were: “In general, I am happy with myself” and “Sometimes, I feel like I am a failure.” A total self-esteem score was calculated by adding up the scores on all 10 items. Higher scores corresponded with higher levels of self-esteem whereas lower scores corresponded with lower levels of self-esteem. The internal consistency of the scale was high with a Cronbach’s α of 0.76 (De Corte et al., 2007). For this sample, α was 0.750.

Children’s Self-Perception of Academic Competence

It was assessed with the Self-Perception Profile for Children (Harter, 1985; translated by Veerman et al., 2004). This questionnaire measures how children perceive their own competences on several life domains. For the current study, self-perceived competence on the school level was used. The total score for that scale was calculated by adding up children’s scores on six questions. For each question, the child had to choose between two sentences and then indicate if that sentence is somewhat or entirely true for them. Every item received a score ranging from 1 to 4. The internal consistency was good, with a Cronbach’s α of 0.78 (Veerman et al., 2004). For the current sample, α for self-perceived academic competence was 0.809.

Mathematical Abilities

As outcome measures, fact retrieval speed and procedural accuracy were investigated. To measure the *fact retrieval speed*, the Arithmetic Number Fact Test (TTR; de Vos, 2002) was used. Children had to solve as much additions (e.g., “7 + 2”), subtractions (e.g., “6 – 5”), multiplications (e.g., “5 × 8”), divisions (e.g., “27 : 9”), or a mix of these exercises as possible in 5 min. The number of correct answers was used as outcome measure. This test has been standardized for Flanders on a sample of 10,059 children (Ghesquière and Ruijsenaars, 1994). The

psychometric value of the test has been demonstrated with a Cronbach’s alpha of 0.90 (Desoete and Roeyers, 2005). For this sample Cronbach’s α was 0.954.

To measure the *procedural accuracy skills* of the child, the Cognitive Developmental skills in aRithmetics Test (CDR; Desoete and Roeyers, 2002) was administered. This test evaluates the understanding and proficiency needed to solve 90 exercises in a number-problem or word-problem format (e.g., “283 times more than –71 is . . .”; “27681 : 90 = . . .”; “Wim has 4.8 kg of flour. Jan has a double amount of flour. How many flour do Jan and Wim have together?”) without a time limit. The number of correct answers was calculated as outcome measure. The CDR has been standardized on 1332 Flemish children (Desoete and Roeyers, 2005). The internal consistency for this sample was Cronbach’s $\alpha = 0.860$.

Statistical Analyses

Before conducting statistical analyses to examine the hypotheses, the missing data (2.381% empty cells) was examined to assess if these items were missing completely at random (MCAR). Little’s MCAR test confirmed that data was missing completely at random, $\chi^2(68, n = 114) = 63.569, p = 0.630$. Missing values were imputed with the expectation-maximization technique.

Since the assumptions for parametric testing were met, Multivariate Analyses of Covariance (MANCOVA) were conducted on the A and P factors separately to examine the first hypothesis. Intelligence was used as covariate, since the MLD and control group significantly differed on IQ (see section “Sample”). To examine the second and third hypothesis, linear regression analyses were conducted with the A, O, and P factors as predictors for fact retrieval speed on the one hand and as predictors for procedural accuracy on the other hand. Interaction terms with group (MLD or control) were added for those variables of which the MANCOVA (Hypothesis 1) revealed that they differed between both groups. The raw scores of the mathematical tests were transformed into z-scores for each grade separately. This was done by standardizing them by the group means per grade, to correct for age effects. Finally, mediation analyses were conducted to test whether the effect of the A predictors on mathematical performance was mediated by the O and/or P predictors (Hypothesis 4).

RESULTS

Hypothesis 1: There Will Be Differences in Antecedent and Propensity Indicators Between Children With and Without Mathematical Learning Disabilities (MLD)

A MANCOVA was conducted on the A predictors, with MLD status as independent variable and intelligence as covariate. Multivariate results revealed significant differences in A factors, $F(4,112) = 21.738, p < 0.001, \eta_p^2 = 0.437$. Furthermore, on the P factors, a similar MANCOVA was conducted.

Multivariate results revealed significant differences in P factors, $F(12,104) = 7.760, p < 0.001, \eta_p^2 = 0.472$. Univariate

results, means (M) and standard deviations (SD s) for the A and P predictors can be found in **Tables 1, 2**, respectively.

Parents of children in the MLD group had significantly lower aspirations [antecedent (A)]. Additionally, these children scored significantly lower on openness, conscientiousness, emotional stability, autonomous motivation, self-esteem, self-perceived competence, intelligence, and working memory when compared to children in the control group. In contrast, they scored significantly higher for negative affect and BAS (P).

Hypothesis 2: A Selection of Antecedent, Opportunity, and Propensity Indicators Will Predict Fact Retrieval Speed

Multiple regression analysis with the A variables, group (MLD or control) and interaction of parental aspirations $\times$ group as predictors for the z -scores on the TTR revealed a significant regression equation, $F(6,114) = 17.256, p < 0.001, R^2 = 0.483$. The regression coefficients, standard deviations and the significance tests for the different predictors can be found in **Table 3A**.

Table 3A demonstrates a significant main effect for parental aspirations and group (MLD or control) on fact retrieval scores. The interaction effect of parental aspirations $\times$ group was also significant (see left part of **Figure 3**). The regression line for the MLD group was non-significant, $F(1,60) = 3.290, p = 0.075, R^2 = 0.051$. However, parental aspirations were predictive for fact retrieval speed in the control group, $F(1,52) = 33.163, p < 0.001, R^2 = 0.385$.

Next, the multivariate regression with the O predictors for fact retrieval was significant, $F(2,114) = 4.079, p < 0.001, R^2 = 0.066$. The univariate results and coefficients can be found in **Table 3B** and indicated that the years of experience the teacher had was predictive for fact retrieval speed.

Further, the multiple regression analysis with the P variables, group (MLD or control) and separate interaction variables (Group $\times$ BAS, $\times$ openness, $\times$ conscientiousness, $\times$ emotional stability, $\times$ autonomous motivation, $\times$ self-esteem, $\times$ negative affect, $\times$ self-perceived competence, $\times$ intelligence, and $\times$ working memory) as predictors was conducted on fact retrieval speed. This analysis revealed a significant regression equation, $F(24,114) = 4.281, p < 0.001, R^2 = 0.525$ (see **Table 3C**).

Results showed a significant main effect of intelligence and a trend for group on fact retrieval speed. Furthermore, the interaction of self-perceived competence $\times$ group was significant (see right part of **Figure 3**). The regression lines

per group indicated that self-perceived competence was a significant predictor for fact retrieval speed in the control group, $F(1,52) = 24.295, p < 0.001, R^2 = 0.314$, but not in the MLD group $F(1,60) = 0.711, p = 0.402, R^2 = 0.012$.

Hypothesis 3: A Selection of Antecedent, Opportunity, and Propensity Indicators Will Predict Procedural Accuracy

Linear regression analyses were conducted for the third hypothesis to predict the z -scores on the CDR. The same interaction variables as in Hypothesis 2 were added into the model.

The multiple regression analysis with the A variables, group (MLD or control) and interaction of parental aspirations $\times$ group as predictors for procedural accuracy revealed a significant regression equation, $F(6,114) = 19.819, p < 0.001, R^2 = 0.517$ (see **Table 4A**).

There was a significant main effect of SES on procedural accuracy and a trend towards significance for parental aspirations as predictor. Next, the multivariate regression with the O factors as predictors for procedural accuracy was significant, $F(2,114) = 3.898, p < 0.001, R^2 = 0.063$. The univariate results and coefficients can be found in **Table 4B**. None of the predictors seemed to be predictive on the univariate level, however, the years of experience of the teacher was marginally significant.

A multiple regression analysis with the P variables, group (MLD or control) and separate interaction variables (Group $\times$ BAS, $\times$ openness, $\times$ conscientiousness, $\times$ emotional stability, $\times$ autonomous motivation, $\times$ self-esteem, $\times$ negative affect, $\times$ self-perceived competence, $\times$ intelligence, and $\times$ working memory) as predictors was conducted on procedural accuracy. This analysis revealed a significant regression equation, $F(24,114) = 8.770, p < 0.001, R^2 = 0.694$ (see **Table 4C**).

The univariate results indicated a significant main effect for positive affect and intelligence on procedural accuracy. Furthermore, there was a trend towards a significant main effect for negative affect, emotional stability, and conscientiousness. The interaction effects for group $\times$ working memory and group $\times$ self-perceived competence were significant (see **Figure 4**). Working memory was a significant predictor for procedural accuracy in the control group, $F(1,52) = 26.117, p < 0.001, R^2 = 0.330$, but not in the MLD group $F(1,60) = 2.025, p = 0.160, R^2 = 0.032$. Also for self-perceived competence, a

TABLE 1 | Multivariate Analyses of Covariance (MANCOVA) on Antecedent predictors with intelligence as covariate.

	MLD		Control		F	p	η_p^2
	M	SD	M	SD			
Parental Aspirations	63.923	8.769	82.668	8.909	86.352	<0.001***	0.429
SES	42.913	10.512	44.021	10.717	0.703	0.403	0.006
Birth Order	1.715	0.906	1.632	0.821	0.460	0.499	0.004
Birth Weight	3284.504	469.444	3288.083	421.967	0.024	0.877	0.000

η_p^2 interpretation: 0.020 = small effect; 0.130 = medium effect; 0.260 = large effect; * $p < 0.050$; ** $p \leq 0.010$; *** $p \leq 0.001$; MLD = Mathematical Learning Disabilities.

TABLE 2 | Multivariate Analyses of Covariance (MANCOVA) on Propensity predictors with intelligence as covariate.

	MLD		Control		<i>F</i>	<i>p</i>	η_p^2
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>			
BIS	2.827	0.539	2.642	0.494	2.671	0.105	0.023
BAS	3.117	0.475	2.974	0.399	4.131	0.044*	0.035
Openness	85.095	11.067	94.780	10.244	8.459	0.004**	0.069
Conscientiousness	99.238	20.331	106.543	17.191	8.125	0.005**	0.066
Emotional Stability	45.175	11.007	51.854	10.352	4.501	0.036*	0.038
Autonomous Motivation	2.893	0.864	3.334	0.933	4.380	0.039*	0.037
Controlled Motivation	2.929	0.732	2.780	0.826	0.000	0.997	0.000
Working Memory	15.762	3.306	21.291	3.804	34.700	<0.001***	0.232
Positive Affect	3.544	0.625	3.606	0.623	0.118	0.731	0.001
Negative Affect	2.380	0.587	2.091	0.555	4.295	0.040*	0.036
Total Self-Esteem	19.714	4.567	22.200	3.638	5.392	0.022*	0.045
Self-Perceived Competence	12.762	3.609	18.324	3.644	43.636	<0.001***	0.275

η_p^2 interpretation: 0.020 = small effect; 0.130 = medium effect; 0.260 = large effect; * $p < 0.050$; ** $p \leq 0.010$; *** $p \leq 0.001$; MLD = Mathematical Learning Disabilities.

significant regression equation was found in the control group, $F(1,54) = 37.647$, $p < 0.001$, $R^2 = 0.415$, but not in the MLD group, $F(1,62) = 1.380$, $p = 0.245$, $R^2 = 0.022$. There was a trend towards a significant effect for the interaction between group $\times$ self-esteem on procedural accuracy.

Hypothesis 4: The Predictive Value of Some Antecedent Variables Will Be Mediated Through Some Opportunity and Propensity Variables

Mediation analyses were conducted (Hypotheses 2 and 3) in “Process” by Andrew Hayes (Field, 2016). Since parental aspirations $\times$ group (MLD or control) was a significant predictor for fact retrieval speed, it was examined if this effect was mediated through teachers’ experience (significant O predictor) on the one hand and through self-perceived competence $\times$ group and intelligence (significant P predictors) on the other hand.

The results revealed that the effect of parental aspirations on fact retrieval speed was not mediated through teachers’ experience for both the MLD and the control group. Results for the MLD group are $b = 0.000$, BCa CI $[-0.007, 0.004]$ and for the control group $b = 0.000$, BCa CI $[-0.007, 0.008]$.

Further, a significant indirect effect of parental aspirations on fact retrieval speed through self-perceived competence was revealed for both the MLD group, $b = 0.004$, BCa CI $[0.000, 0.011]$, and the control group, $b = 0.013$, BCa CI $[0.002, 0.028]$. The indirect effect of parental aspirations through intelligence was non-significant, $b = 0.004$, BCa CI $[-0.002, 0.012]$.

Because SES was a significant predictor for procedural accuracy, a possible mediation through intelligence and positive affect on the one hand and self-perceived competence $\times$ group (significant P predictors) on the other hand was examined. Mediation through intelligence was significant, $b = 0.014$, BCa CI $[0.005, 0.024]$. No indirect effect was found for SES on procedural accuracy through positive affect, $b = -0.002$, BCa CI $[-0.006, 0.001]$. The indirect effect of SES on procedural accuracy through

self-perceived competence was non-significant for both the MLD group, $b = -0.000$, BCa CI $[-0.007, 0.007]$ and the control group, $b = 0.005$, BCa CI $[-0.001, 0.014]$.

DISCUSSION

Throughout the last decade, several predictors of mathematical learning have been proposed. To evaluate individual differences and the unique contribution of predictors, it is important to take into account the interrelationships between those predictors. Within the O-P Model, it is suggested that learning occurs as the result of A, O, and P factors (Byrnes and Miller, 2007). Studies on large secondary datasets have revealed the value of this model in kindergarten (Byrnes and Wasik, 2009; Wang and Byrnes, 2013), the beginning of primary school (Byrnes and Wasik, 2009), and in secondary school (Byrnes and Miller, 2007, 2016). To the best of our knowledge, no studies have examined this model by collecting primary data in the second half of primary school in a group of children with and without MLD. This study aimed to fill this gap in the existing research by investigating whether children with and without MLD differ on A and P variables and assess whether information on the combination of these variables adds to the current knowledge on mathematical abilities and disabilities. Moreover as mathematics has been described as componential in nature (Dowker, 2015), the relationship with both fact retrieval speed and procedural calculation is examined and compared.

Differences in Antecedent and Propensity Indicators Between Children With and Without MLD

Results showed significant differences in both A and P factors when comparing children with and without MLD.

In contrast with the hypotheses, children with MLD did not differ significantly from typically developing children on the A factors, with regards to birth weight, SES, and

TABLE 3 | Multivariate Regression Models with Antecedent, Opportunity, or Propensity predictors on fact retrieval speed (tested with the TTR).

		B	SE	β	<i>t</i>	<i>p</i>
(A) Antecedent predictors $F = 17.256, R^2 = 0.483 (p < 0.001^{***})$	(constant)	1.548	0.917		1.687	0.094
	Parental Aspirations	−0.024	0.011	−0.313	−2.215	0.029*
	SES	−0.001	0.006	−0.006	−0.080	0.936
	Birth Order	−0.123	0.080	−0.109	−1.539	0.127
	Birth Weight	−9.849 ^a −005	0.000	−0.045	−0.648	0.518
	Group (MLD, Control)	−5.450	1.184	−2.790	−4.603	<0.001***
	Group × Parental Aspirations	0.085	0.016	3.623	5.357	<0.001***
(B) Opportunity predictors $F = 4.079, R^2 = 0.066 (p = 0.019^*)$	(constant)	−0.820	0.358		−2.292	0.024
	Teachers' Experience (years)	0.020	0.010	0.185	2.019	0.046*
	Hours of math per week	0.108	0.066	0.149	1.622	0.108
(C) Propensity predictors $F = 4.281, R^2 = 0.525 (p < 0.001^{***})$	(constant)	1.271	1.686		0.754	0.453
	BIS	−0.072	0.163	−0.039	−0.444	0.658
	BAS	0.181	0.254	0.082	0.715	0.477
	Openness	0.016	0.011	0.196	1.443	0.155
	Conscientiousness	−0.009	0.006	−0.173	−1.393	0.167
	Emotional Stability	−0.007	0.012	−0.075	−0.541	0.590
	Autonomous Motivation	0.006	0.130	0.005	0.043	0.966
	Controlled Motivation	0.094	0.103	0.075	0.919	0.360
	Working Memory	0.035	0.050	0.396	0.702	0.484
	Intelligence	−0.030	0.011	−0.405	−2.672	0.009**
	Positive Affect	−0.168	0.146	−0.107	−1.157	0.250
	Negative Affect	−0.092	0.195	−0.056	−0.474	0.637
	Total Self-Esteem	0.041	0.026	0.180	1.549	0.125
	Self-Perceived Competence	−0.040	0.034	−0.188	−1.178	0.242
	Group (MLD, Control)	−4.519	2.480	−2.313	−1.822	0.072
	Group × BAS	0.176	0.386	0.272	0.455	0.650
	Group × Openness	−0.020	0.018	−0.986	−1.136	0.259
	Group × Conscientiousness	0.013	0.010	0.742	1.343	0.182
	Group × Emotional Stability	0.011	0.017	0.311	0.663	0.509
	Group × Autonomous Motivation	−0.056	0.175	−0.101	−0.318	0.751
	Group × Working Memory	0.035	0.050	0.396	0.702	0.484
	Group × Intelligence	0.029	0.016	1.584	1.897	0.061
	Group × Negative Affect	0.002	0.296	0.002	0.006	0.995
	Group × Total Self-Esteem	−0.063	0.044	−0.729	−1.409	0.162
	Group × Self-Perceived Competence	0.166	0.050	1.611	3.317	0.001**

* $p < 0.050$; ** $p \leq 0.010$; *** $p \leq 0.001$; MLD = Mathematical Learning Disabilities.

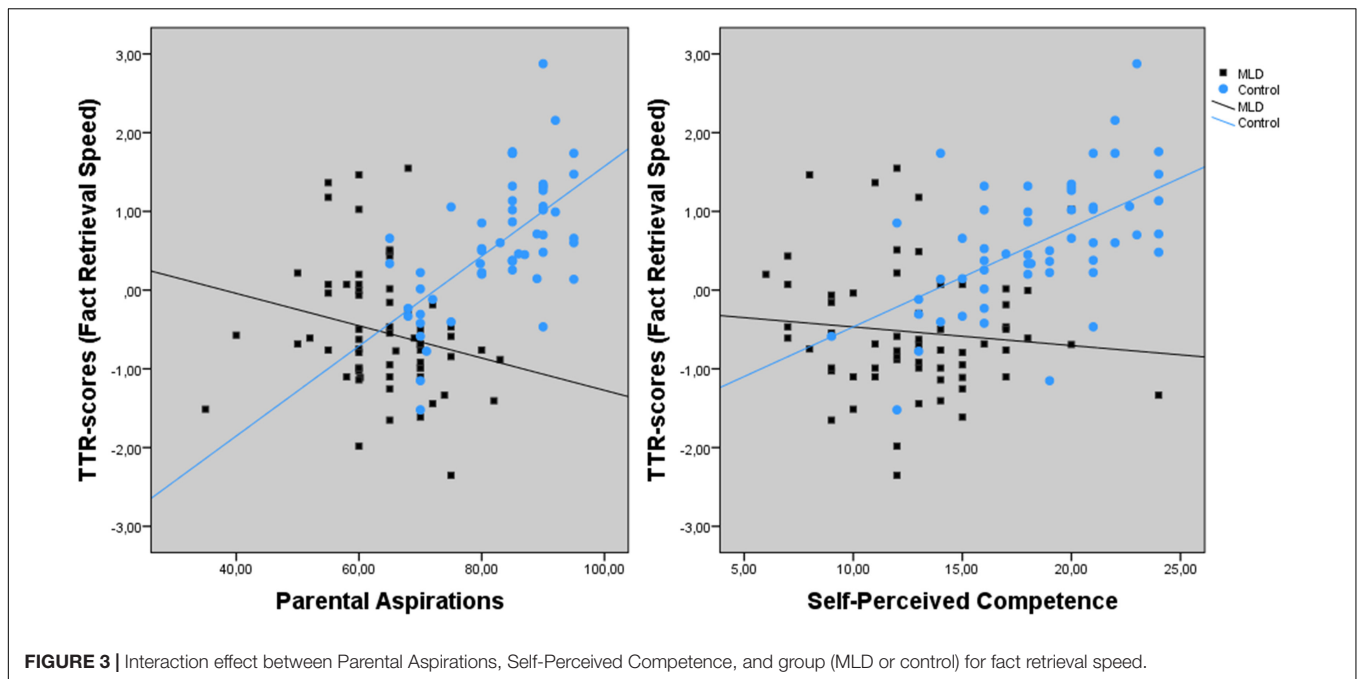
birth order. However, they did differ on parental aspirations. Parents had significantly lower aspirations toward mathematical learning when their children had MLD. A large effect size was found. Further research is needed to examine whether these lower aspirations were caused by children's continuous struggle with math learning or if the lower math performances of children with MLD are due to lower parental aspirations.

With regards to the P factors, children with and without MLD differed on temperament, personality, motivation, working memory, SWB, self-esteem, and self-perceived competence, after controlling for intelligence. Regarding temperament, children with MLD had higher scores on BAS compared to typically developing children. However, in contrast with Van Beek et al. (2013), results did not show significant differences between both groups for BIS. This unexpected result could be due to a power

problem. Our findings seem to indicate that children with MLD might be more sensitive for rewards than peers without MLD. This might implicate that teachers should use rewards and positive consequences as a lever to enhance their mathematical performances.

Concerning personality, children with MLD were less open to new experiences, were less conscientious, and had lower scores for emotional stability compared to peers in the control group. The effects of openness and conscientiousness were larger than the effect of emotional stability (versus neuroticism). These results are in line with earlier research which indicated openness and conscientiousness as the personality traits most associated with mathematical performance (Poropat, 2009; Zhang and Ziegler, 2016).

Analysis of the P factor of motivation indicated no differences in the amount of controlled motivation (where the force to fulfill



a task is external; e.g., a reward) between children with and without MLD. However, children with MLD had lower levels of autonomous motivation (where one fulfills a task for an internal reason such as passion or future relevance of the topic) when compared to controls. This indicates that children from both groups were equally motivated for mathematics because they had to, whereas children with MLD were less motivated for mathematics because they wanted to. Next, in line with literature (Roth et al., 2015; Peng and Fuchs, 2016), results revealed that children with MLD experienced more difficulties with working memory when compared to typically developing children. This effect was between medium and large indicating that working memory problems might be impactful for children with MLD.

Furthermore, children with MLD experienced more negative affect on a regular school day than their typically developing peers in the same school context. There were no significant group differences found for positive affect. When examining self-esteem, data revealed that children with MLD reported lower self-esteem than their peers without MLD. These results indicate the impact of MLD on the SWB of children. Even though they seemed to experience the same amount of positive feelings as their typically developing peers, they experienced more negative affect and more negative feelings toward themselves. In line with the reciprocal-effects model (Guay et al., 2003; Seaton et al., 2015), it is possible that having MLD impacts children's SWB, which in its turn affects their mathematical abilities resulting in more severe math problems.

Finally, children with MLD perceived their own academic competences much lower (large effect size) than did typically developing children, which indicated that they were aware of their own lower capacity in mathematics.

The Predictive Value of Antecedent, Opportunity, and Propensity Factors for Math Performance

First, some A, O, and P factors were predictive for fact retrieval speed. The combination of SES, birth weight, parental aspirations, and birth order as A predictors explained 48.3% of variance in fact retrieval speed. In line with earlier research (Blevins-Knabe et al., 2007; Kleemans et al., 2012; Niklas et al., 2016), parental aspirations towards mathematical performance were a significant A predictor. Parents who wanted their children to score higher at the end of the current school year tended to have children who performed better in mathematics. Nonetheless, in our dataset, parental aspirations were important predictors only for typically developing children. Additionally, this effect was partially mediated through children's self-perceived competence. However, based on the current study, no conclusions can be drawn about the direction of the effect. For instance, it is possible that lower math abilities of children influenced parental aspirations and children's self-perceived competence. In contrast, it is possible that lower parental aspirations influenced children's self-perceived competence and in their turn resulted in lower math abilities. However, reciprocal effects are also a possibility. Additional and longitudinal studies are necessary to understand the effect of parental aspirations more clearly. Moreover, not finding a predictive effect of parental aspirations for fact retrieval speed in the MLD group could be associated with severity or specificity of MLD as a developmental disorder. It is possible that persevering fact retrieval or fluency problems that characterize MLD cannot be influenced by parents' expectations.

In contrast with the available literature on A predictors, SES, birth weight, and birth order did not significantly predict fact

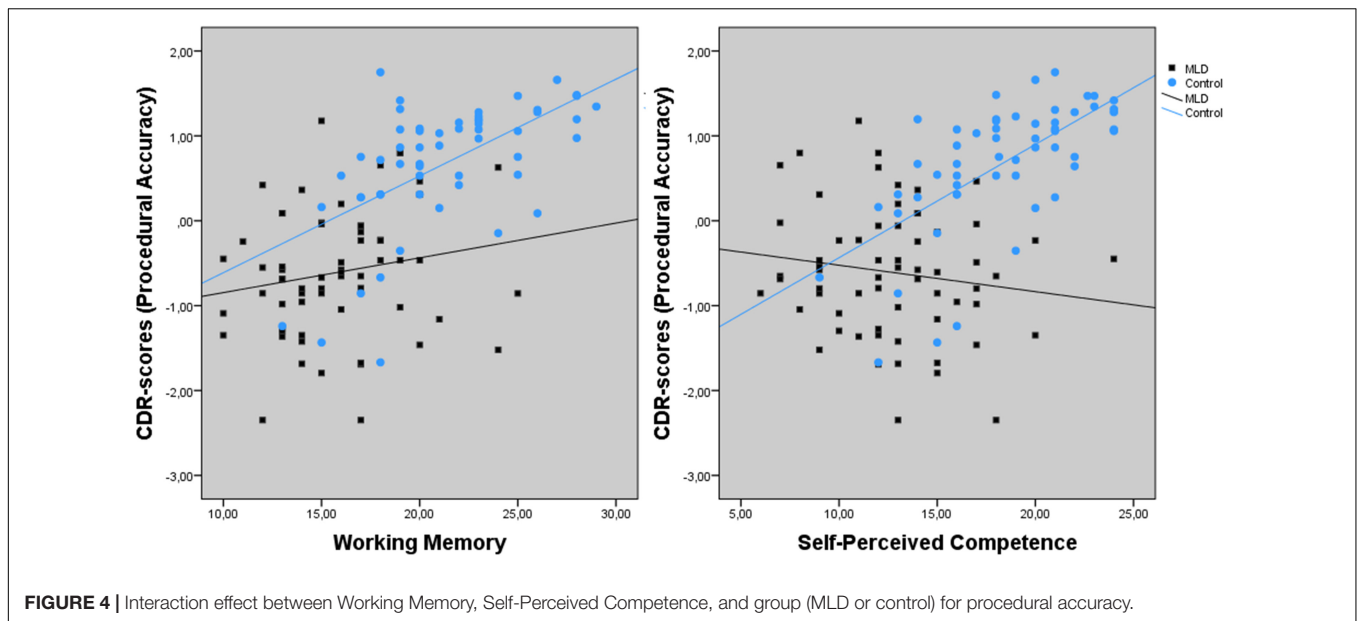
TABLE 4 | Multivariate Regression Models with Antecedent, Opportunity, or Propensity predictors on procedural accuracy (tested with the CDR).

		B	SE	β	<i>t</i>	<i>p</i>
(A) Antecedent predictors $F = 19.819, R^2 = 0.517 (p < 0.001^{***})$	(constant)	-2.133	0.896		-2.358	0.020
	Parental Aspirations	0.017	0.011	0.227	1.659	0.100
	SES	0.014	0.006	0.146	2.171	0.032*
	Birth Order	0.080	0.078	0.070	1.027	0.307
	Birth Weight	0.000	0.000	-0.046	-0.686	0.494
	Group	-0.486	1.156	-0.246	-0.420	0.675
	Group $\times$ Parental Aspirations	0.017	0.015	0.735	1.125	0.263
(B) Opportunity predictors $F = 3.898, R^2 = 0.063 (p = 0.023^*)$	(constant)	-0.812	0.362		-2.243	0.027
	Teachers' Experience (years)	0.019	0.010	0.180	1.965	0.052
	Hours of math per week	0.107	0.067	0.146	1.596	0.113
(C) Propensity predictors $F = 8.770 R^2 = 0.694 (p < 0.001^{***})$	(constant)	-3.834	1.369		-2.800	0.006
	BIS	-0.029	0.132	-0.016	-0.223	0.824
	BAS	0.203	0.206	0.091	0.983	0.328
	Openness	0.000	0.009	0.002	0.018	0.986
	Conscientiousness	-0.013	0.008	-0.720	-1.623	0.108
	Emotional Stability	-0.020	0.014	-0.548	-1.453	0.150
	Autonomous Motivation	0.034	0.105	0.031	0.319	0.750
	Controlled Motivation	-0.045	0.083	-0.035	-0.534	0.594
	Working Memory	-0.002	0.028	-0.008	-0.065	0.949
	Intelligence	0.023	0.009	0.317	2.608	0.011*
	Positive Affect	-0.323	0.118	-0.203	-2.731	0.008**
	Negative Affect	0.302	0.158	0.179	1.904	0.060
	Total Self-Esteem	-0.031	0.021	-0.134	-1.434	0.155
	Self-Perceived Competence	-0.015	0.028	-0.068	-0.533	0.595
	Group (MLD, Control)	-0.645	2.014	-0.326	-0.320	0.750
	Group $\times$ BAS	-0.105	0.313	-0.160	-0.334	0.739
	Group $\times$ Openness	0.014	0.014	0.694	0.995	0.322
	Group $\times$ Conscientiousness	-0.013	0.008	-0.720	-1.623	0.108
	Group $\times$ Emotional Stability	-0.020	0.014	-0.548	-1.453	0.150
	Group $\times$ Autonomous Motivation	-0.223	0.142	-0.402	-1.568	0.120
	Group $\times$ Working Memory	0.082	0.041	0.908	2.006	0.048*
	Group $\times$ Intelligence	-0.017	0.013	-0.886	-1.322	0.190
	Group $\times$ Negative Affect	-0.038	0.240	-0.043	-0.159	0.874
	Group $\times$ Total Self-Esteem	0.071	0.036	0.819	1.973	0.052
	Group $\times$ Self-Perceived Competence	0.128	0.041	1.225	3.141	0.002**

* $p < 0.050$; ** $p \leq 0.010$; *** $p \leq 0.001$; MLD = Mathematical Learning Disabilities.

retrieval speed. The lack of association between SES and fact retrieval speed might be explained by the limited sample size or by the nature of fact retrieval mathematics as component of mathematics. Since retrieving arithmetic facts depends on drill and memorization, it could be less susceptible to the job and educational level of parents than other components of mathematics. On birth weight, the literature focused especially on effects of extremely low birth weight (<1500 g; De Rodrigues et al., 2006; Chatterji et al., 2014). In the current sample, none of the children had birth weights below 2000 g. Further, the results of this study did not confirm earlier studies which reported better performance in academic contexts when higher in the birth order (Hotz and Pantano, 2015). This might be due to the small variability in birth order places of the participants, since 86.4% of this sample was the first or second born child in their family.

This study confirmed that mathematical abilities improve with more O's (Zhang, 2008; Byrnes and Wasik, 2009). The O's explained 6.6% of the variance in fact retrieval speed. More experienced teachers seem to have a positive impact on children's math performances. The number of hours of mathematics instruction children received per week had no significant effect. This might indicate that not the quantity (number of hours) but the quality (teachers' experience) of instruction matters. Furthermore, mediation of A through O variables was not found in the current study since parental aspirations did not predict teachers' experience. This might be explained by the specific selection of O variables in the current study compared to earlier studies on the O-P Model (e.g., Byrnes and Wasik, 2009) which included richer O measurements. Future research should measure O factors more broadly, whereas now teachers were only asked about



their years of experience and how many hours they taught mathematics.

The P variables included in this study explained 52.5% of the variance found in fact retrieval speed, which indicated that the P factors are the strongest predictors for retrieving arithmetic facts. This is in line with earlier research on the O-P Model (Byrnes and Wasik, 2009). Both intelligence and self-perceived competence were significant predictors in earlier studies (Arefi et al., 2014; Peng and Fuchs, 2016). However, the effect of self-perceived competence was only present in typically developing children, but not in children with MLD. Analog to parental aspirations, this might be related to the severity of fact retrieval deficits and might not be influenced by having higher perceptions of your own competences. It is important to note that also here, no conclusions can be drawn about the direction of the effects. Longitudinal studies are necessary but in line with the literature we can expect reciprocal effects between academic self-concept and academic achievement (Guay et al., 2003; Seaton et al., 2015).

Second, procedural accuracy could be predicted by some of the A, O, and P factors. The combination of SES, birth weight, parental aspirations, and birth order as A predictors explained 51.7% of the variance in procedural accuracy. Children with higher SES, performed better in procedural calculation, which is in line with earlier research (Wang et al., 2013). The data on parental aspirations of this study did not confirm its predictive value for procedural calculation, in contrast with the existing literature (Blevins-Knabe et al., 2007; Kleemans et al., 2012; Niklas et al., 2016). However, this could be related to power-issues since there was a trend towards a significant effect.

Analysis of procedural calculation, confirmed that mathematical abilities become better with more O's (Zhang, 2008; Byrnes and Wasik, 2009). O's explained 6.3% of the variance. There was a trend towards a significant

association for the years of experience the teacher had. The same conclusions could be drawn as for fact retrieval fluency.

The P variables were the most predictive for procedural calculation, which is in line with earlier research on the O-P Model (Byrnes and Wasik, 2009). They explained 69.4% of variance. Intelligence, positive affect, working memory, and self-perceived competence were significant predictors. Higher levels of intelligence were associated with higher scores in procedural accuracy, which is in line with the literature (Roth et al., 2015; Peng and Fuchs, 2016). Moreover, the effect of SES on procedural calculation abilities was partially mediated through intelligence in this sample. With regards to working memory capacity, a significant association with procedural accuracy was found in the typically developing children. This association is in line with work of De Weerd et al. (2013). A positive association between self-perceived competence, and procedural calculation was found, confirming results from earlier studies (Arefi et al., 2014). However, the effect of self-perceived competence was only present in typically developing children, not in children with MLD. This is analog to the findings on self-perceived competence and fact retrieval speed. Again, it is reasonable to expect reciprocal effects between self-perceived competence and procedural accuracy (Guay et al., 2003; Seaton et al., 2015). Not finding an effect of self-perceived competence for children with MLD might be the result of severe deficits that are not susceptible for influences of self-perceived competence. Additionally, a negative association was found between positive affect and procedural accuracy, which is in contrast with the literature on SWB (Quinn and Duckworth, 2007). Additional research is necessary to confirm and explain this finding.

Third, when comparing the effects of A, O and, P variables on fact retrieval speed with procedural accuracy, some important similarities and differences should be noted.

Antecedent (A) factors explained about half of the variance in both types of math learning. However, for fact retrieval, the most important predictor was parental aspirations, whereas for procedural accuracy, SES seemed more important than parental aspirations. Furthermore, results revealed that the impact of A variables was mediated through P variables. More specifically, the effect of parental aspirations on fact retrieval speed was partially mediated through children's self-perceived competence and the effect of SES on procedural accuracy was partially mediated through intelligence. These results provide evidence for the structure of the O-P Model (Byrnes and Miller, 2007). Nonetheless, in contrast with the proposed structure of the model, the data of this study did not confirm the mediation of A variables through O factors, which could be explained by the rather limited measures of O variables in the current study. Future research should include richer measurements of O's. For O variables, about 6% of the variance for each of the mathematical components could be explained. Teachers' years of experience proved to be an important factor, which highlights the importance of the quality and not quantity of instruction. The P variables were the strongest predictors of math abilities and more variance could be explained for procedural accuracy (about 70%) compared to fact retrieval fluency (about 50%). In earlier research on the O-P Model (Byrnes and Miller, 2007), P variables were also the strongest predictors for outcome. However, in the current study different P's seemed to be predictive for fact retrieval compared to procedural calculation. Intelligence and self-perceived competence contributed to both types of mathematics, whereas positive affect and working memory were only predictive for procedural calculation. This could be related to the nature of the tasks used. In procedural calculation, children have to understand the mathematical principles and procedures to find the correct answer. Compared to fact retrieval tasks where arithmetic facts have to be memorized and retrieved, it makes sense that procedural accuracy is more susceptible to other influences than intelligence (e.g., positive affect and working memory). Moreover, fact retrieval depends on drill and memorization and therefore retrieving arithmetical facts might be less susceptible to the influence of P variables in general. This could also be the explanation of why more variance is explained by P variables for procedural calculation compared to fact retrieval fluency.

Finally, in contrast with the hypotheses, no association with mathematical abilities was found for motivation, personality and temperament. Although the literature on personality describes conscientiousness and openness as the most predictive personality traits for academic performances (Poropat, 2009; Zhang and Ziegler, 2016), when these variables were simultaneously investigated with other P variables in a holistic model, no predictive value was revealed. This emphasizes the importance of taking into account the interrelationship between several predictors in order to thoroughly understand mathematical development. However, we did find significant differences in personality when comparing children with and without MLD (see section "Differences in Antecedent and Propensity Indicators Between Children With and

Without MLD"). Regarding autonomous motivation and temperament factors (BIS and BAS), this study did not reveal a predictive value for mathematical abilities when investigated within a holistic model. This is in contrast with the existing literature on motivation (Taylor et al., 2014) and temperament. Nonetheless, we did find significant differences for these variables when comparing children in the MLD group with their typically developing peers (see section "Differences in Antecedent and Propensity Indicators Between Children With and Without MLD"). When trying to predict outcome, it seems to be important to examine multiple variables within a holistic framework and to compare children with and without MLD.

Limitations and Suggestions for Future Research

Every study has limitations. In this study, the sample size was rather small which could have repercussions on results. The sample size in previous work on the O-P Model was much larger. However, the data used in this study were primary collected data from an MLD population, whereas all previous studies used secondary data from a general population. We should take into account that some significant associations or differences on population level could not be detected within this sample due to power issues but it is a strength that data is collected within a clinical population. Future research should collect primary data on larger sample sizes.

Furthermore, in previous studies on the O-P Model, prior knowledge was a strong predictor of math performance (Byrnes and Miller, 2007). In the current study, this variable could not be examined since we were not able to collect data that was comparable across children. The children lived in different cities and attended different schools. Additionally, there were no standardized measures of prior knowledge previously administered in all children. However, in a follow-up study the collected measures of current skills will be used as prior knowledge for their skills in wave 2.

Finally, because this was a cross-sectional study, no conclusions about cause-and-effect can be made. Additional, longitudinal studies are currently being conducted.

CONCLUSION AND IMPLICATIONS FOR PRACTICE

Despite the limitations, our results support the fact that children with MLD differ on A (e.g., parental aspirations) as well as on several (both cognitive and non-cognitive) P indicators. An exclusive P approach or only assessing cognitive predictors might not be a good idea.

Second, the O-P Model revealed to be applicable to the study of children with MLD. However, our findings also demonstrated that general protocols for the assessment of procedural calculation abilities or fact retrieval speed should not be implemented in the same way to test children with MLD and their typically developing peers. Since different predictors for

mathematical abilities were found in children with and without MLD and in line with the componential nature of mathematics, adequately customized and broad assessments remain needed. Regarding procedural calculation, our findings revealed the importance of questionnaires on SES and tests on intelligence, positive affect, working memory, self-esteem, and self-perceived competence. With regards to fact retrieval speed, questionnaires on parental aspirations, and teachers' experience, intelligence tests and a questionnaire on self-perceived competence seem indicated.

Finally, the current findings seem to indicate that children with MLD might be more sensitive to rewards, less open to new experiences and less conscientious. In addition, they were less autonomously motivated and had lower levels of SWB, lower self-esteem and lower self-perceived competence. These findings suggest the importance of positive feedback and psychoeducation including the enhancement of the autonomous motivation for mathematics in those children, in addition to the focus on their math acquisition. Therapy should focus on

their strengths and reward small positive steps in the correct direction.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of "The Ethical Committee of the Faculty of Psychology and Educational Sciences of Ghent University" with written informed consent from the parents of all subjects. All parents gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

EB performed the data collection, data analysis, and writing of the manuscript. AD supervised the data analysis and wrote the manuscript.

REFERENCES

- Arefi, M., Naghibzadeh, M., and Boloki, A. (2014). The relationship of parental attachment, peer attachment, and academic self-concept to academic achievement of high school students. *Int. J. Acad. Res.* 6, 73–78. doi: 10.1016/j.jsp.2012.10.001
- Aunio, P., and Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learn. Individ. Dif.* 20, 427–435. doi: 10.1016/j.lindif.2010.06.003
- Baten, E., Praet, M., and Desoete, A. (2017). The relevance and efficacy of metacognition for instructional design in the domain of mathematics. *Int. J. Math. Educ.* 49, 613–623. doi: 10.1007/s11858-017-0851-y
- Blevins-Knabe, B., Whiteside-Mansell, L., and Selig, J. (2007). Parenting and mathematical development. *Acad. Exch. Q.* 11, 76–80.
- Bryant, D. P., Bryant, B. R., and Pfannenstiel, K. H. (2015). Mathematics interventions: translating research into practice mathematics interventions: translating research into practice. *Interv. School Clin.* 50, 255–256. doi: 10.1177/1053451214560893
- Byrnes, J. P., and Miller, D. (2016). The growth of mathematics and reading skills in segregated and diverse schools: an opportunity-propensity analysis of a national database. *Contemp. Educ. Psychol.* 46, 34–51. doi: 10.1016/j.cedpsych.2016.04.002
- Byrnes, J. P., and Miller, D. C. (2007). The relative importance of predictors of math and science achievement: an opportunity – propensity analysis. *Contemp. Educ. Psychol.* 32, 599–629. doi: 10.1016/j.cedpsych.2006.09.002
- Byrnes, J. P., and Wasik, B. A. (2009). Factors predictive of mathematics achievement in kindergarten, first and third grades: an opportunity – propensity analysis. *Contemp. Educ. Psychol.* 34, 167–183. doi: 10.1016/j.cedpsych.2009.01.002
- Carver, C. S., and White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: the BIS/BAS Scales. *J. Pers. Soc. Psychol.* 67, 319–333. doi: 10.1037/0022-3514.67.2.319
- Cattaneo, M. A., Oggenfuss, C., and Wolter, S. C. (2016). The more, the better? the impact of instructional time on student performance. *Educ. Econ.* 25, 433–445. doi: 10.1080/09645292.2017.1315055
- Ceulemans, A., Baten, E., Loey, T., Hoppenbrouwers, K., Rousseau, S., and Desoete, A. (2017). The relative importance of parental numerical opportunities, prerequisite knowledge and parent involvement as predictors for early math achievement in young children. *Interdiscip. Educ. Psychol.* Available at: <http://riverapublications.com/article/the-relative-importance-of-parental-numerical-opportunities-prerequisite-knowledge-and-parent-involvement-as-predictors-for-early-math-achievement-in-article-info>
- Chatterji, P., Dohyung, K., and Kajal, L. (2014). Birth weight and academic achievement in childhood. *Health Econ.* 23, 1013–1035. doi: 10.1002/hec.3074
- Clements, D. H., and Sarama, J. (2011). Early childhood mathematics intervention. *Science* 333, 968–970. doi: 10.1126/science.1204537
- Cohen Kadosh, R., and Dowker, A. (eds). (2015). *The Oxford Handbook of Numerical Cognition*. Oxford: Oxford University Press.
- Costa, P. T. J., and McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Cowan, R. (2015). "Education," in *The Oxford Handbook of Numerical Cognition*, eds R. Cohen Kadosh and A. Dowker (Oxford: Oxford University Press), 1021–1035.
- De Corte, K., Buysse, A., Verhofstadt, L. L., Roeyers, H., Ponnet, K., and Davis, M. H. (2007). Measuring empathic tendencies: reliability and validity of the Dutch version of the interpersonal reactivity index. *Psychol. Belg.* 47, 235–260. doi: 10.5334/pb-47-4-235
- De Rodrigues, M. C. C., Mello, R. R., and Fonseca, S. C. (2006). Learning difficulties in schoolchildren born with very low birth weight. *J. Pediatr.* 82, 6–14. doi: 10.2223/JPED.1429
- de Vos, T. (2002). *Tempo Test Rekenen (TTR)*. Amsterdam: Harcourt.
- De Weerd, F., Desoete, A., and Roeyers, H. (2013). WM in children with RD and/or MD working memory in children with reading disabilities and/or mathematical disabilities. *J. Learn. Disabil.* 46, 461–472. doi: 10.1177/0022219412455238
- Desoete, A. (2008). Co-Morbidity in mathematical learning disabilities: rule or exception? *Open Rehabil. J.* 1, 15–26. doi: 10.2174/1874943700801010015
- Desoete, A., and Roeyers, H. (2002). *Cognitieve Deelvaardigheden Rekenen*. Herentals: Vlaamse Vereniging voor Logopedisten.
- Desoete, A., and Roeyers, H. (2005). Cognitive skills in mathematical problem solving in Grade 3. *Br. J. Educ. Psychol.* 75, 119–138. doi: 10.1348/000709904X22287
- Dix, A., and van der Meer, E. (2015). Arithmetic and algebraic problem solving and resource allocation: the distinct impact of fluid and numerical intelligence. *Psychophysiology* 52, 544–554. doi: 10.1111/psyp.12367
- Dowker, A. (2015). "Individual differences in arithmetical abilities. The componential nature of arithmetic," in *The Oxford Handbook of Mathematical Cognition*, eds R. C. Kadosh and A. Dowker (Oxford: Medicine UK), 862–878.
- Egberink, I. J. L., Holly-Middelkamp, F. R., and Vermeulen, C. S. M. (2010). *COTAN Beoordeling, Hiërarchische Persoonlijkheidsvragenlijst voor Kinderen*. Available at: www.cotandocumentatie.nl
- Engelen, U., De Peuter, S., Victoir, A., Van Diest, I., and Van den Bergh, O. (2006). Further validation of the positive and negative affect schedule (PANAS) and comparison of two Dutch versions. *Gedrag Gezondh* 34, 89–102.

- Field, A. (2016). *Discovering Statistics using IBM SPSS*. Thousand Oaks, CA: SAGE Publications.
- Franck, E., De Raedt, R., Barbez, C., and Rosseel, Y. (2008). Psychometric properties of the Dutch Rosenberg self-esteem scale. *Psychol. Belg.* 48, 25–35. doi: 10.5334/pb-48-1-25
- Franken, I. H. A., Muris, P., and Rassin, E. (2005). Psychometric properties of the Dutch BIS/BAS Scales. *J. Psychopathol. Behav. Assess.* 27, 25–30. doi: 10.1007/s10862-005-3262-2
- Froiland, J. M., and Worrell, F. C. (2016). Intrinsic motivation, learning goals, engagement, and achievement in a diverse high school. *Psychol. Schools* 53, 321–336. doi: 10.1002/pits.21901
- Ghesquière, P. (2014). “Actualisering van het standpunt in verband met de praktijk van attesteren voor kinderen met een leerstoornis in het gewoon onderwijs,” in *Zorg Dragen voor Kinderen en Jongeren met Leerproblemen. Handvaten voor Goede Praktijk*, eds P. H. Ghesquière, A. Desoete, and C. Andries (Leuven: Acco), 11–19.
- Ghesquière, P., and Ruijsenaars, A. (1994). *Vlaamse Normen voor Studietoetsen Rekenen en Technisch Lezen lager Onderwijs*. Leuven: KUL-CSBO.
- Gray, J. A. (1981). “A critique of Eysenck’s theory of personality,” in *A Model for Personality*, ed. H. J. Eysenck (New York, NY: Springer), 246–276.
- Grégoire, J. (2000). *L’évaluation clinique de l’intelligence de l’enfant: Théorie et pratique du WISC-III*. Liège: Mardaga.
- Guay, F., Marsh, H. W., and Boivin, M. (2003). Academic self-concept and academic achievement: developmental perspectives on their causal ordering. *J. Educ. Psychol.* 95, 124–136. doi: 10.1037/0022-0663.95.1.124
- Harter, S. (1985). *Self Perception Profile for Children*. Colorado, AZ: University of Denver.
- Henik, A., Rubinstein, O., and Ashkenazi, S. (2015). “Developmental dyscalculia as a heterogeneous disability,” in *The Oxford Handbook of Numerical Cognition*, eds R. Cohen Kadosh and A. Dowker (Oxford: University Press), 662–677.
- Hotz, V. J., and Pantano, J. (2015). Strategic parenting, birth order, and school performance. *J. Popul. Econ.* 28, 911–936. doi: 10.1007/s00148-015-0542-3
- Jordan, N. C., and Levine, S. C. (2009). Socioeconomic variation, number competence, and mathematics learning difficulties in young children. *Dev. Disabil. Res. Rev.* 15, 60–68. doi: 10.1002/ddrr.46
- Kleemans, T., Peeters, M., Segers, E., and Verhoeven, L. (2012). Child and home predictors of early numeracy skills in kindergarten. *Early Child. Res. Q.* 27, 471–477. doi: 10.1016/j.ecresq.2011.12.004
- Kort, W., Schittekatte, M., Bosmans, M., Compaan, E. L., Dekker, P. H., Vermeir, G., et al. (2005). *WISC-III-NL Wechsler Intelligence Scale For Children, Derde Editie NL. Handleiding en Verantwoording*. Amsterdam: Harcourt Test Publishers/Nederlands Instituut voor Psychologen.
- Kort, W., Schittekatte, M., and Compaan, E. (2008). *CELF-4-NL: Clinical Evaluation of Language Fundamentals - vierde editie*. Amsterdam: Pearson Assessment and Information B.V.
- Kucian, K., and von Aster, M. (2015). Developmental dyscalculia. *Eur. J. Pediatr.* 174, 1–13. doi: 10.1007/s00431-014-2455-7
- Mervielde, I., and de Fruyt, F. (2009). *Hiërarchische Persoonlijkheidsvragenlijst voor Kinderen (HiPIC)*. Amsterdam: Hogrefe.
- Mooij, T., and Driessen, G. (2008). Differential ability and attainment in language and arithmetic of Dutch primary school pupils. *Br. J. Educ. Psychol.* 78, 491–506. doi: 10.1348/000709907X235981
- Murayama, K., Pekrun, R., Suzuki, M., Marsh, H. W., and Lichtenfeld, S. (2016). Don’t aim too high for your kids: parental overaspiration undermines students’ learning in mathematics. *J. Pers. Soc. Psychol.* 111, 766–779. doi: 10.1037/pspp0000079
- Niklas, F., Cohrsen, C., and Tayler, C. (2016). Improving preschoolers’ numerical abilities by enhancing the home numeracy environment. *Early Educ. Dev.* 27, 372–383. doi: 10.1080/10409289.2015.1076676
- Peng, P., and Fuchs, D. (2016). A meta-analysis of working memory deficits in children with learning difficulties: is there a difference between verbal domain and numerical domain? *J. Learn. Disabil.* 49, 3–20. doi: 10.1177/0022219414521667
- Pieters, S., Roeyers, H., Rosseel, Y., Waelvelde, H., Van, and Desoete, A. (2015). Identifying subtypes among children with developmental coordination disorder and mathematical learning disabilities, using model-based clustering. *J. Learn. Disabil.* 48, 83–95. doi: 10.1177/0022219413491288
- Pieters, S., Van Vreckem, C., Vanderswalmen, R., Desoete, A., and Van Waelvelde, H. (2009). DCD + leerstoornis: één + één is meer dan twee. *Signaal* 22, 28–35.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychol. Bull.* 135, 322–338. doi: 10.1037/a0014996
- Quinn, P. D., and Duckworth, A. L. (2007). “Happiness and academic achievement: evidence for reciprocal causality,” in *Poster Presented at the Annual Meeting of the American Psychological Society*, Washington, DC.
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F., and Spinath, F. M. (2015). Intelligence and school grades: a meta-analysis. *Intelligence* 53, 118–137. doi: 10.1016/j.intell.2015.09.002
- Ryan, R., and Deci, E. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol.* 55, 68–78. doi: 10.1037/0003-066X.55.1.68
- Savelsbergh, E. R., Prins, G. T., Rietbergen, C., Fechner, S., Vaessen, B. E., Draijer, J. M., et al. (2016). Effects of innovative science and mathematics teaching on student attitudes and achievement: a meta-analytic study. *Educ. Res. Rev.* 19, 158–172. doi: 10.1016/j.edurev.2016.07.003
- Scheiris, J., and Desoete, A. (2008). De prevalentie van enkele specifieke ontwikkelings- en gedragsstoornissen en hun comorbiditeit. *Signaal* 62, 4–14.
- Schuchart, C., Buch, S., and Piel, S. (2015). Characteristics of mathematical tasks and social class-related achievement differences among primary school children. *Int. J. Educ. Res.* 70, 1–15. doi: 10.1016/j.ijer.2014.12.002
- Seaton, M., Marsh, H. W., Parker, P. D., Craven, R. G., and Yeung, A. S. (2015). The reciprocal effects model revisited. *Gift. Child Q.* 59, 143–156. doi: 10.1177/0016986215583870
- Shalev, R. S. (2007). *Why is Math So Hard for Some Children? The Nature and Origins of Mathematical Learning Difficulties and Disabilities*, eds D. B. Berch and M. M. M. Mazzocco (Baltimore, MD: Paul H. Brookes Publishing), 49–60.
- Shin, M., and Bryant, D. P. (2015). A synthesis of mathematical and cognitive performances of students with mathematics learning disabilities: a synthesis of mathematical and cognitive performances of students with mathematics learning disabilities. *J. Learn. Disabil.* 48, 96–112. doi: 10.1177/0022219413508324
- Siemann, J., and Petermann, F. (2018). Evaluation of the triple code model of numerical processing—reviewing past neuroimaging and clinical findings. *Res. Dev. Disabil.* 72, 106–117. doi: 10.1016/j.ridd.2017.11.001
- Smits, D. J. M., and Boeck, P. D. (2006). From BIS / BAS to the big five. *Eur. J. Pers.* 270, 255–270. doi: 10.1002/per.583
- Stock, P., Desoete, A., and Roeyers, H. (2010). Detecting children with arithmetic disabilities from kindergarten: evidence from a 3-year longitudinal study on the role of preparatory arithmetic abilities. *J. Learn. Disabil.* 43, 250–268. doi: 10.1177/0022219409345011
- Taylor, G., Jungert, T., Mageau, G. A., Schattke, K., Dedic, H., Rosenfield, S., et al. (2014). A self-determination theory approach to predicting school achievement over time: the unique role of intrinsic motivation. *Contemp. Educ. Psychol.* 39, 342–358. doi: 10.1016/j.cedpsych.2014.08.002
- Van Beek, I., Kranenburg, I. C., Taris, T. W., and Schaufeli, W. B. (2013). BIS- and BAS-activation and study outcomes: a mediation study. *Pers. Individ. Dif.* 55, 474–479. doi: 10.1016/j.paid.2013.04.013
- Vanbinst, K., Ghesquière, P., and De Smedt, B. (2015). Does numerical processing uniquely predict first graders’ future development of single-digit arithmetic? *Learn. Individ. Dif.* 37, 153–160. doi: 10.1016/j.lindif.2014.12.004
- Vansteenkiste, M., Sierens, E., Soenens, B., Luyckx, K., and Lens, W. (2009). Motivational profiles from a self-determination perspective: the quality of motivation matters. *J. Educ. Psychol.* 101, 671–688. doi: 10.1037/a0015083
- Veerman, J., Straathof, M., Treffers, P., Van den Bergh, B., and ten Brink, L. (2004). *Competentiebelevingsschaal voor kinderen (CBSK)*. Lisse: Harcourt.
- Wang, A., and Byrnes, J. P. (2013). Does the opportunity – propensity framework predict the early mathematics skills of low-income pre-kindergarten children? *Educ. Psychol.* 38, 259–270. doi: 10.1016/j.cedpsych.2013.04.004
- Wang, A. H., Shen, F., and Byrnes, J. P. (2013). Does the opportunity-propensity framework predict the early mathematics skills of low-income pre-kindergarten children? *Contemp. Educ. Psychol.* 38, 259–270. doi: 10.1016/j.cedpsych.2013.04.004

- Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070. doi: 10.1037/0022-3514.54.6.1063
- Zhang, D. (2008). *The Effects of Teacher Education Level, Teaching Experience, and Teaching Behaviors on Student Science Achievement*. Ph.D. thesis, Utah State University, Logan, UT.
- Zhang, J., and Ziegler, M. (2016). How do the big five influence scholastic performance? A big five-narrow traits model or a double mediation model. *Learn. Individ. Dif.* 50, 93–102. doi: 10.1016/j.lindif.2016.07.011

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

Copyright © 2018 Baten and Desoete. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Response-To-Intervention in Finland and the United States: Mathematics Learning Support as an Example

Piia M. Björn^{1*}, Mikko Aro², Tuire Koponen², Lynn S. Fuchs³ and Douglas Fuchs³

¹ School of Educational Sciences and Psychology, University of Eastern Finland, Joensuu, Finland, ² Department of Education, University of Jyväskylä, Jyväskylä, Finland, ³ Department of Special Education, Peabody College, Vanderbilt University, Nashville, TN, United States

Response to Intervention (RTI) was accepted in the early 2000s as a new framework for identifying learning difficulties (LD) in the U.S. In Finland, a similar multi-tiered framework has existed since 2010. In the present study, these frameworks are presented from the viewpoint of the role of assessment and instruction as expressed in documents that describe the frameworks, as it seems that these two components of RTI are the most disparate between the U.S. and Finland. We present a suggestion for the Finnish framework as an example of support in mathematics learning that incorporates principles of RTI (such as systematized assessment and instruction, cyclic support, and modifiable instruction). Finally, recommendations are presented for further refining and developing assessment and instruction policies in the two countries.

OPEN ACCESS

Edited by:

Annamie Desoete,
Ghent University, Belgium

Reviewed by:

Lara Ragpot,
University of Johannesburg,
South Africa
Celestino Rodríguez,
Universidad de Oviedo Mieres, Spain

*Correspondence:

Piia M. Björn
piia.bjorn@uef.fi

Specialty section:

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

Received: 23 February 2018

Accepted: 04 May 2018

Published: 05 June 2018

Citation:

Björn PM, Aro M, Koponen T,
Fuchs LS and Fuchs D (2018)
Response-To-Intervention in Finland
and the United States: Mathematics
Learning Support as an Example.
Front. Psychol. 9:800.
doi: 10.3389/fpsyg.2018.00800

Keywords: comparative study, response to intervention framework, assessment, instruction, support in mathematics

Why do we need educational frameworks and guidelines for providing support? Why can teachers not rely on their education and knowledge of learning and provide sufficient instruction and support for all students in need of something extra? These are the questions we discuss in the present paper. Different countries have different approaches to these matters, but we choose to compare the multi-tiered frameworks for support in learning used in the United States and Finland, as interesting similarities and dissimilarities exist. In the U.S., Response-To-Intervention (RTI) has long been a suggested framework for identifying students with disabilities. It provides guidelines for early prevention and for delivering evidence-based instruction with intensifying tiers of support. Close monitoring of student progress is also at the core of the U.S. RTI. Informed decision making at all levels within the system (administrative, teacher, and parental; see Fuchs and Fuchs, 2005) is provided. The basic idea of RTI in the U.S. is that the school provides the child with research-based instruction while the child is in the general education environment, and the school adjusts the intensity or nature of assessment and instruction according to the student's progress (Fuchs and Fuchs, 2005).

In our previous paper on the U.S. RTI and Finnish "RTI" (Björn et al., 2016), we found that the original purpose and, subsequently, the definition of RTI framework in these countries differed to some extent. The present paper on assessment and instruction within RTI frameworks in the U.S. and Finland is an extension of the previous papers. We previously found that RTI in the U.S. was primarily developed for LD (Learning Difficulty) identification, and the Finnish version was primarily intended to re-structure the existing support service framework for struggling students (Björn et al., 2016). Instead of the Finnish framework, the prevention of LD was an acknowledged goal in the frameworks of both countries. It seemed that the two frameworks were similar in appearance but differed in content and delivery. We wanted more knowledge that would explain why the renewed

Finnish framework was outlined similar to the U.S. RTI, however, we found that the massive amount of existing knowledge on the pros and cons of the approach seemed to be neglected by the formal documents defining Finnish version of the framework, as many important definitions were not made explicit in formal documents.

We further realized that the role of the special education service system differed within the RTI framework in different parts of the U.S., while in Finland, special educational services have the same role within an RTI-like framework throughout the country (Björn et al., 2016). Thus, the present work presents the frameworks in the two countries but with a specific focus on assessment and instruction. The goal is to determine ways to refine both frameworks, with a special emphasis on bringing forward support for mathematics learning in Finland. We start this paper by briefly introducing the creation and implementation process of RTI in the U.S. and Finland. This is followed by differences and similarities in the definitions of assessment and instruction in these countries at each tier of support. Then, a suggestion for structuring support in mathematics in Finland is presented. We conclude by discussing possibilities for further refinements of the RTI approach in both countries.

THE CREATION AND IMPLEMENTATION OF RTI IN THE U.S. AND FINLAND

Some earlier studies have examined the differences between identification and learning support frameworks in the U.S. and Finland (see Ikonen and Jahnukainen, 2007; Jahnukainen, 2011; Björn et al., 2016). However, information about distinctions in the ways these countries operationalize RTI assessment and instruction is absent. The types of policy papers that present and compare educational frameworks implemented in different countries are important because even though the processes behind the reforms differ, the actual need for constructing frameworks for support in learning stems from the same source. That is, all education systems try to teach students effectively and at a reasonable cost. Such reforms are also nationwide processes, and each country may learn something from other countries despite cultural differences.

The U.S. school system consists of public and private schools. The average school age ranges from about age 5–18 years. The Finnish school system is public; there are basically no private schools. Children enter the compulsory schooling system the year they turn seven years old. Compulsory schooling lasts nine years, until the child reaches the age of 15 or 16 (depending on the time of year the child was born). The overall educational standards are run by the Ministry of Education, Science, and Culture, but the schools may relatively freely implement support in their own curricula (www.minedu.fi).

Although RTI in the U.S. as an approach to identifying and instructing especially students with LD has a long history (dating back to the 1970's), the implementation of RTI after the Individuals with Disabilities Education Act, (IDEA, 2004), enacted in 2004, has been interpreted as somewhat problematic.

For example, Zirkel and Thomas (2010) conducted a survey that addressed the early years of RTI implementation in the U.S. Those authors found that although RTI has been an allowable substitute for the widely used IQ discrepancy criteria since 2004 (see Fuchs and Fuchs, 2006), confusion still persists between, for example, legal requirements and professional recommendations. Zirkel and Thomas have concluded that the legal content of RTI is still somewhat incomplete. This probably explains why countless versions of RTI have emerged. However, schools in the U.S. may still use the IQ discrepancy model along with RTI (see Zirkel, 2012a,b,c) in the process of identifying LD. Although RTI models vary considerably from state to state and from district to district in the U.S., many approaches are comparable to the three-tiered RTI framework currently in use in Finland (Björn et al., 2016).

In Finland, the Ministry of Education, Science, and Culture formed a steering group in 2006 to focus on developing a strategy for special education in basic education. Several tasks were to be achieved: developing ways to analyze the need for the amount of special educational services, developing legislation concerning special education, developing teacher education, developing administrative procedures in special educational services, and developing other areas related to special education. Consequently, a new strategy for special education was published in 2007.

Based on this strategy document, a renewed Basic Education Act was introduced in 2010 and was officially implemented in August 2011 in all Finnish schools (Pesonen et al., 2015). This led to a framework with three levels of support for learning: Tier 1 general support (including co-teaching, differentiated teaching, etc. as forms of support); Tier 2 intensified support (domain-specific learning plans and support in reading, writing in flexible groups in addition to the forms of support mentioned before); and Tier 3 special support (all previous forms of support and individualized education plans) at each level, the student is entitled to a variety of forms of support (e.g., even special education, see Björn et al., 2016).

RTI as an approach to the identification and support of LD is gradually being implemented throughout Europe. For example, in the Netherlands, the Dutch Act on “Passend Onderwijs” adopted in 2014, states that all children should be included in mainstream education as much as possible, with financial support provided to schools by regional educational administrations. In addition to this, there is growing interest in using this framework throughout many countries in primary education (Scholvinck and Janssen, 2014). According to the interpretation of RTI in the Netherlands, Tier 1 support is provided inside the classroom by the classroom teacher. This includes direct and differentiated instruction for all students. However, Tier 2 and Tier 3 support is mostly provided by a remedial teacher outside the classroom.

The U.S. RTI system has two main approaches to instruction: the problem-solving model and the standard protocol model (Fuchs et al., 2010; Jenkins et al., 2013). In the problem-solving model, a student's deficits are addressed by implementing a research-based intervention specially designed for that individual student (Johnson et al., 2006; Fuchs et al., 2010). Typically in the problem-solving model, decision-making teams, which may consist of teachers, administrators, school psychologists,

and parents, follow a recursive four-step process: (a) define the problem, (b) plan an intervention, (c) implement the intervention, and (d) evaluate the student's progress (Fuchs et al., 2003; Bender and Shores, 2007). In the standard protocol model, students with similar difficulties (e.g., problems with reading fluency) are given research-based interventions that have been standardized and proven effective for students with similar difficulties for a predetermined amount of time (Johnson et al., 2006). The problem-solving model resembles the Finnish framework more than the standard protocol approach (Björn et al., 2016).

RTI ASSESSMENT AND INSTRUCTION

Because LDs (typically in reading or math or both) are major reasons for the need of extra support in learning (Fletcher et al., 2007), relevant guidelines for both assessment and instruction are needed (Fuchs et al., 2010, 2012). The concept of assessment is often viewed as unidirectional. It used to be interpreted as an authority administering assessments, with the examinee viewed as an object of classification (Ysseldyke et al., 1983). In RTI, however, as Grigorenko (2009) has noted, the roots of assessment in RTI seem to be related to dynamic assessment (DA; see also Elliott, 2000, 2003; Fuchs et al., 2007, 2011) in which the assessment is flexibly intertwined with teaching sequences. This enables up-to-date assessment results that can quickly inform the instruction. Relevant and supplementary skills-based testing is also an important component of RTI assessment as is progress monitoring. It has been proposed that the performance of “nonresponders,” (i.e., those children who do not show progress in academic skills) is monitored frequently with a set of short instruments relevant to these skills (Fuchs and Fuchs, 2005). By monitoring a student's learning and comparing it to that of peers receiving the same instruction, teachers can determine whether the student's academic level and rate of progress warrant further assessment or formal evaluation (Fuchs and Fuchs, 2005).

The first important assumption acknowledged in both RTI and DA is that conventional assessment does not work for children who have diverse educational and cultural experiences. These children are often those who need more intensified support in learning. The second assumption is that, instead of focusing on children's skills and abilities at a specific time (Fuchs et al., 2010), children have the potential to learn with adequate education or intervention (Fuchs et al., 2007). The third assumption is that the reason for assessment is to inform intervention, and consequently, the results of assessment should have direct implications for selecting or modifying instruction. The assessment data and continuous progress monitoring inform instruction at each tier. Additionally, research-based curriculum and instruction, as well as the systematic assessment of the fidelity with which instruction and interventions are implemented, are essential (National Association of State Directors of Special Education, 2005; Fuchs et al., 2007). It is important to note that assessment also includes other foci than learning outcomes in which the student's task-motivation (Eccles, 2005), academic self-efficacy, and metacognitive skills (Seaton et al., 2013) are taken

into account in addition to the important assessment of the learning environment (Johnson et al., 2006).

Next, we will go through assessment and instruction policies in each overall Tier (the 3-tier RTI frameworks used here) comparing the U.S. and Finland. After that, we will present a model for providing individual support in mathematics according to Finnish RTI framework and legislation.

TIER 1 ASSESSMENT AND INSTRUCTION

See **Table 1** for a comparative presentation of assessment and instruction practices within RTI frameworks in the U.S. and Finland. Tier 1 in the U.S. RTI includes statewide norms as well as suggested materials and assessments usually performed within general education settings. On Tier 1, according to Fuchs and Fuchs (2005), struggling children are identified through poor performance in classwide, schoolwide, or districtwide screening intended to designate which children are at risk of academic or behavioral problems. In Finland, to date, there is *no formal guidance* on performing screenings within the RTI framework. Some type of universal screening might (once or twice per year), however, be performed according to a school's and municipality's own system. Finnish teachers may freely decide when, how, and with which the screenings are performed. The frequency of screening is normally three times per year in RTI, but once again, it is not clearly localized within the Finnish framework.

The latest addition to the screening procedure in the U.S. RTI framework was suggested by Fuchs et al. (2012). Originally, support when moving from Tier 1 to Tier 2 was based on one screening phase according to which students who did not respond to instruction were referred for more intensive support. The new procedure involves a second stage of screening performed after a short period of support, which can contribute to accurate identification of students who require a supplemental layer of reading intervention (Compton et al., 2012) or math intervention (Fuchs et al., 2011). Another innovation by researchers actively working with the U.S. RTI was a second stage of diagnostic assessment that could be used to move students who did not respond to a supplemental layer of tutoring immediately to a more intensive and perhaps long-term intervention they required (Compton et al., 2012). Without such a second stage of screening, schools would provide costly intervention to many students who did not need it. Compton et al. (2012) have suggested a multistage screening process near the beginning of the first grade to avoid an “RTI wait-to-fail” model, in which children are required to participate in 10–30 weeks of supplemental intervention that could have been predicted to be inadequate.

In Finland, an *optional learning plan* is suggested (e.g., in the Basic Education Act, 2010) at the Tier 1 level called “general support.” This plan entails a means for assessment and support. The U.S. version of RTI suggests no such documentation. The frequency of progress monitoring (although it shows significant variation) is high within RTI and is not defined within the Finnish framework. In other words, in the renewed Finnish framework of support in learning, the role of assessment and instruction is

TABLE 1 | Assessment and Instruction on each Tier of RTI/Level of support, Finnish framework.

	RTI	Finnish framework
	Tier 1 (Primary prevention)	General support
ASSESSMENT		
Type of assessment/identification	Universal Screening, statewide norms ^a	Not Specified (NS) ^b
Frequency of screening	3 times per year	NS
Who does assessments	General education teacher	NS
Who makes decisions	Multiprofessional team, parents	Multiprofessional team, parents
Materials used in progress monitoring	CBM	NS, usually group assessments
Type of progress monitoring	Academic skills development monitoring	NS
Frequency of progress monitoring	Weekly, for 5-8 weeks for at-risk students	1-2 times per year
INSTRUCTION		
Length of Tier	max 8 weeks (1 year [^])	NS
Intensity of intervention(s)	90 min daily (in reading)	Within regular school work, NS
Type of interventions	Core, With	NS/With, PT, SG, Ind
Type of instruction	Explicit, top-down (Differentiated instruction)	Differentiated instruction, etc.
Methods of interventions/instruction	Research-principled instruction, curricular [^]	Flexible
Movement criteria between Tiers	Final status	NS
	Tier 2 (Secondary prevention)	Intensified support
ASSESSMENT		
Type of assessment/identification	Instruction-based, skill-specific	NS
Who does assessments	Trained school personnel	School personnel
Who makes decisions	Multiprofessional teams, parents	Multiprofessional teams, parents
Materials used for assessments	Progress monitoring ^c	NS ^c
Type of progress monitoring	CBM	Learning plan assessment
Frequency of progress monitoring	No less than once every 2 weeks [^]	NS
INSTRUCTION		
Length of Tier	max 1 school year (9-30 weeks [^])	NS
Intensity of intervention(s)	min 3 times/week, min 20-30 min/session [^]	"More intense"
Type of interventions	Targeted/SG (3-5 students)	NS/PT, FT, With, SG, Ind
Type of instruction	Standard protocol, replicable (Team problem-solving, Behavioral consultation)	Flexible, NS
Methods of interventions/instruction	Specified programs, scripted protocols, evidence-b.	NS
Movement criteria between Tiers	Final status; cut point slope [∞]	NS
	Tier 3 (Tertiary prevention)	Special support
ASSESSMENT		
Type of assessment/identification	Curriculum-based, diagnostic	NS
Who does assessments	Highly skilled/educated school personnel	School personnel, consultation (medica
Who makes decisions	Multiprofessional teams, parents	Multiprofessional teams, parents
Materials used for assessments	Progress monitoring, diagnostic tools	Standardized tests available, but NS
Type of progress monitoring	CBM, diagnostic tests, IEP	Pedagogic plan assessment, IEP
Frequency of progress monitoring	No less than once a week [^]	NS
INSTRUCTION		
Length of Tier	Min. 15-20 weeks [^]	NS
Intensity of intervention(s)	More frequently than Tier 2, min 30 min/session [^]	"More intense"
Type of interventions	Intense, SG, Ind (1-2 students)	Flexible, Ind, NS
Type of instruction	Data-based instruction (expert consultation)	PT, FT, With, SG, Ind
Methods of interventions/instructionSpecified programs, individual	Specified programs, individual	NS

(Continued)

TABLE 1 | Continued

	RTI	Finnish framework
	Tier 3 (Tertiary prevention)	Special support
Movement criteria between Tiers	Final status; cut point slope, individual progress [^]	Re-assessments especially in transitions

^aScreening, see: <http://www.rti4success.org/screeningTools/>

^b<http://www.lukimat.fi/lukimat-oppimisen-arviointi/materiaalit/tuen-tarpeen-tunnistaminen>: materials for performing universal screening exist but they are not formally linked to the renewed framework.

^cProgress monitoring, see: <http://www.rti4success.org/progressMonitoringTools>, <https://charts.intensiveintervention.org/chart/progress-monitoring>; Finnish progress monitoring would <http://www.lukimat.fi/lukimat-oppimisen-arviointi/materiaalit/oppimisen-seuranta> exist but they are not formally linked to the renewed framework; [^]Johnson, E., Mellard, D., Fuchs, D., McKnight, M. for NRCLD (2006); NS, Not Specified; PT, Part-time special education (in the USA: inclusive teaching); FT, Full-time special education (such as special classes, self-contained classrooms); With, Student within mainstream education, although has LD; SG, Small-group instruction (such as "Tier time," resource rooms), Ind, Individual instruction. ∞ as in performance below/above 25th percentile. £ These examples from New York State Special Education Department website: <http://www.p12.nysed.gov/specialed/RTI/guidance/instruction.htm>

somewhat undefined although the framework mentions possible forms of support (such as co-teaching, smaller study groups, etc.).

According to Fuchs and Fuchs (2005), in the three-tier U.S. RTI model, Tier 1 concerns at-risk children who have been identified through a screening process. They receive research-based instruction, sometimes in small groups, sometimes as part of a classwide intervention. A certain amount of time (generally not more than 6–8 weeks) is allotted to see if the child responds to the instruction. Each student's progress is monitored closely (for more information, see: <http://www.rtinetwork.org/essential/assessment/progress/validated-forms-progressmonitoring>). The intervention programs may be selected from a bank of research-proven interventions based on school resources in the U.S. The concept of progress monitoring (CBM) and a resource bank of suggested intervention methods are not mentioned at all in documents defining the Finnish framework.

TIER 2 ASSESSMENT AND INSTRUCTION

In the U.S. RTI, Tier 2 (also referred to as secondary prevention) belongs to general education as an instructional service. In Finland, this level called "intensified support," including assessment as well as instruction, is organized via consultation and collaboration between teachers. In the U.S. RTI, assessment is instruction-based and skill-specific. The Finnish framework provides no formal guidance for assessment (in the sense of frequency). However, Finnish schools may, for example, decide whether to do a skill-specific assessment of students in need of extra support in learning. The Finnish framework provides for an *obligatory learning plan* at this level of intensified support in which the support a student receives is reported by teachers. No description of frequency or type of progress monitoring exists in the Finnish framework at the level of intensified support. The learning plan document consists of descriptions of different forms of support provided for a student. Large variation exists, as there is no guidance on time for support.

Multi-professional consultation is made in problem-solving RTI frameworks. Evidence-based protocols are used by reading specialists, special education teachers, and paraprofessionals in some RTI versions. Tier 2 within the RTI framework is an important stage between Tier 1 and the intensified Tier 3. Therefore, instruction on Tier 2 is evidence-based as well as

performed in short periods to allow for the instruction to be modified in a timely manner (Fuchs and Fuchs, 2005). According to Fuchs and Fuchs (2005), if the child does not respond to the first level of group-oriented interventions, he or she typically is moved to the next RTI level. The length of time on Tier 2 has been reported to vary between 9 and 30 weeks, even one school year. The time allotted to see if the child responds to interventions at this more intensive level may be longer than on Tier 1. The intervention has been successful if the child shows adequate progress.

The group size of students receiving support given outside classrooms is another important feature on Tier 2 of RTI (Berkeley et al., 2009). For example, the state of Kansas has indicated that small-group instruction should consist of between three and five students on Tier 2 and fewer than three students on Tier 3. Other state models are more flexible in group size requirements. Arizona's model, for example, allows for large- or small-group instruction on Tier 1, small group instruction on Tier 2, and small or individualized instruction on Tier 3.

Within the Finnish framework, small-group instruction, along with the overall instruction that takes into account the diversity of students, is often described as "flexible." This type of support is usually provided by special needs teachers or regular classroom teachers. However, co-teaching is a suggested form of support in the documents that have followed the actual Finnish law (for example, see Ahtiainen et al., 2012).

TIER 3 ASSESSMENT AND INSTRUCTION

Tier 3 in the U.S. framework differs in many ways from the equivalent level of the Finnish framework, which is called "special support." For example, the RTI framework in many US states does not include any form of special education at this tier (although it has been frequently suggested by researchers in the field, see the work of Fuchs and Fuchs, 2005, for example). In contrast, this tier entirely belongs to special education in Finland although a student might still receive support and instruction in regular classroom instruction. If the support offered within the first two RTI tiers in the US has not been enough, significantly more intensified (no less than once a week for 15–20 weeks) instruction is then essential (Fuchs and Fuchs, 2005). Furthermore, if the child does not respond to instruction

at this level, then he or she is likely to be referred for a full and individual evaluation. This referral is a major difference between the U.S. version of RTI and the Finnish framework. The child has already been assessed many times during the level of intensified support in Finland, but not in a unified manner across municipalities or schools within the same district as there is a lack in formal guidance on performing the assessment. Access to special education services in Finland does not require statements of eligibility but is based on multidisciplinary decision-making that also involves the caregivers' opinions. The U.S. RTI provides for instruction for one or two students at a time. The Finnish system lacks explicit min-max descriptions for different levels of support, but many times a student on Tier 3 is situated, at least part of a day, in small groups outside the regular classroom for the most important content areas (usually literacy skills and mathematics). All possible forms of instruction are in use at this level of support in Finland. An obligatory pedagogical review is conducted of all students, and the existing means of support and goals for learning are defined in this review.

Tier 3 RTI in the U.S. has an interesting feature: individualized data-based instruction (or experimental teaching; for a case example, see Fuchs et al., 2010). DBI is a research-based process for individualizing and intensifying interventions through the systematic use of assessment data, validated interventions, and research-based adaptation strategies (see more at: <http://www.intensiveintervention.org>). This form of instruction resembles in many ways the flexibility and degree of individual assessment and instruction that exists in the Finnish framework; teaching methods are individually adjusted. However, what is missing from the background of the Finnish framework is a research-based resource center that would actually validate using individually adjusted instructional methods.

Assessment and instruction in the U.S. RTI framework seem to be closely intertwined. First, the forms of assessment are defined in more detail in the U.S. framework. Second, the main forms of instruction/intervention delivered to students within the U.S. RTI framework rely on research-based interventions, which often include well-defined assessment and programmatic content designed to ensure intensity and duration (Fuchs and Fuchs, 2006). In contrast, the Finnish framework does not include clear definitions for support or follow-up of learning results. Because students with severe learning difficulties in mathematics are in need of the most intensive support, we will next present a suggestion for refining the Finnish framework in terms of individual support in mathematics. Note that our suggestion might be used in other content areas as well.

FINNISH FRAMEWORK FOR INDIVIDUAL SUPPORT IN MATHEMATICS: AN EXAMPLE OF RTI INTERPRETATION

We have identified a national need for bringing more content and research-based substance to the RTI-like framework, as well as a more systematized approach in Finland. This can be done

by providing the support stipulated in formal legislation and other documents schools and teachers currently use in their everyday work. We have not tried to present everything as so much better in the U.S., by using U.S. RTI as an example, but we want to point out that the way Finnish three-tiered framework is currently presented has left too much room for local interpretation. By discussing this in an international forum, we believe that other countries currently in the process of developing their own RTI frameworks might be able to handle building and implementing the framework even better than Finland and the U.S.

We have published a more comprehensive Finnish version of this suggestion on support in mathematics (see Björn et al., 2015) incorporating all three tiers of support, but we have rethought and refined the model in terms of Tier 3 support in mathematics for the present paper. Overall, our suggestion needs to omit some of the principles already in use in the U.S., but that is mainly due to the current lack of material (e.g., assessment tools, progress monitoring tools, etc.) Our suggestion follows Finnish legislation and the outline of the Finnish RTI or "three-tiered framework for support" but incorporates the suggestions of Gersten et al. (2009) and Bryant et al. (2014).

Slavin and Lake (2008) have pointed out that the best learning results in mathematics may be achieved by using systematized, yet flexible, ways of support. Which means that teachers should be given possibilities to modify the support offered (see, Lemons et al., 2014). In Finland, special educational services (as in support provided by a special needs teacher) are available at all three tiers of support. However, the main principle should be that the more intensive the need for support, the more individualized support should be given (Gersten et al., 2009). Consequently, Tier 3 support in primary school should mean choosing evidence-based intervention material as the basis for planning mathematics instruction (Mononen, 2014).

Tier 1 and Tier 2 support precede Tier 3. If preliminary support for learning mathematics in the classroom as part of a large group or even occasionally as part of a small group had been attempted without clear signs of acceleration of math skills, then, according to Finnish law, a formal referral to special education would be needed for Tier 3 support. Subsequently, an individual education plan (IEP) with plans for instruction would be drawn up with the participation of the student, caregivers, school psychologist, classroom teacher, and special needs teacher. We suggest that approach to instruction during a school year would consist of several cycles. The current situation in schools is that each teacher (or teacher and special needs teacher-pair) decide on the frequency and content of support. This results in differing practices, and the rights of suitable instruction provided for each individual student in need of support in mathematics are not addressed adequately.

To correct this situation, we suggest that each cycle of support lasts for 5–7 weeks, and that the support is provided 3–4 times per week (each session duration 30 min of intensive work). starting from making sure very basic math skills are learned (number line skills backwards and onwards, calculations including additions, subtractions, overall estimation ability). By viewing the support as cycles throughout the school year, groups/pairs of students

participating in Tier 3 support could work together regardless of age. We recommend that students work in pairs or small groups, maybe occasionally even provided with fully individualized support. This means that based on “what works” literature (see Gersten et al., 2009), students that need intensive support in mathematics would benefit the most from support given in smaller groups rather than in a large classroom group. We have conducted an intervention with trials on individual support provided in a regular classroom for students in need of support in mathematics Björn et al., (under review), but those inclusion trials did not fully convince us of their superiority to small-group intervention outside the regular classroom. Consequently, we think that intensive intervention periods provided as relatively short cycles, instead of continuous intensive instruction, would enable testing the regular classroom as a learning environment occasionally, and, if sufficient skills have been learned, the “pull out” type of instruction/intervention outside the regular classroom could be stopped at some point.

Each support cycle would begin and end with a short assessment of learning gains so that adjustments of instruction could be made in a timely manner. A cyclic assessment also enables the teacher to determine the point at which each basic math skill has been learned. This way, the approach for assessment and instruction would be “continuous” in terms of what we know about the persistence of developmental mathematics learning difficulties (Fletcher et al., 2007).

We cannot expect severe mathematics learning difficulties to be “cured” even by repeating several cycles of support during a school year; instead, support would need to be provided over several school years. The teaching contents during these support sessions would include basic arithmetic and estimation skills, according to individual needs, for as long as deemed necessary. Continuity of the support would be ensured by keeping a record of the support and assessment given to each student. Giving many alternate suggestions for intervention programs to be selected from as the instructional basis for this support cannot be done at this time. This is due to the fact that, to date, only a few intervention programs for mathematics are available in Finland (for more information, see www.lukimat.fi).

What we have presented here can be summed up like this; assessment and instruction on Tier 3 (special support) should be continuous, cyclic, individual, and based on evidence-based intervention programs. Support can be provided in many different contexts, but it must be systematized and modifiable between cycles.

DISCUSSION

In this paper, we presented the RTI framework and the three-level Finnish educational support system from the viewpoints of assessment and instruction. The models were implemented based on similar background philosophies: the right to receive the best possible preventive support for learning and participation. To the recent Finnish reform (Basic Education Act, 2010), after many phases, developed into a model in similar to the U.S. RTI model (Fuchs and Fuchs, 2005), at least on the surface. However, there are many differences that might give new insights to any

country planning to develop similar frameworks. For example, the current U.S. model aims for the identification and prevention of further learning difficulties (Compton et al., 2012) by placing a student within a suitable tier of intervention (Vaughn and Fuchs, 2003). The Finnish model, in contrast, mainly aims at supporting learning at the earliest time point possible (Opetusministeriö. Erityisopetuksen strategia, 2007) within the three-tiered framework.

The major finding of the present analysis is that unlike the renewed Finnish system of support in education, the U.S. RTI framework included as early as 2004 many suggested materials for universal screening, early intervention, multi-tiered levels of support, evidence-based intervention, data-based decision-making regarding intervention, and using students’ responsiveness to evidence-based instruction in evaluating disability status (Haager et al., 2007). RTI in the U.S. has succeeded in accelerating a paradigmatic change in the uses of testing. Instead of focusing on learning achievement at one point, RTI focuses on individual responses in relation to instruction (Fletcher and Vaughn, 2009; Fuchs et al., 2010).

Moreover, the concept of evidence-based teaching or evidence-based intervention is not present in either of the Finnish documents (Opetusministeriö. Erityisopetuksen strategia, 2007; Basic Education Act, 2010) or in Finnish schools. In the Finnish model, individual assessment (progress when receiving support) is not described. Thus, one major observation that might explain why there is such a noticeable difference between RTI and the Finnish framework is that there is no such large degree of teacher accountability in Finnish school culture (see Sahlberg, 2010) as may be observed to exist within the U.S. For example, the concept of “fidelity to instruction” (Fuchs et al., 2007) is not yet in use in Finland. Instead, the concept of “trust” is used frequently (see Itkonen and Jahnukainen, 2007) when talking about teachers’ work.

Municipalities, schools, and teachers in Finland have a relatively broad autonomy in interpreting legislation and curricular instructions. One reason for this is the equity of the Finnish educational policy system (Linnakylä et al., 2011). Another reason for this type of freedom is that Finnish teachers must have a Master’s degree in education to be recruited to a permanent teaching position. Due to this high educational level, Finnish teachers are often deemed as trusted professionals. Therefore, they are used to making decisions on how to assess students’ skills, what type of instruction to apply, and how long to give instruction before making a decision on whether or not to move the student to the next level of support. This results in very individual and different ways of supporting students’ learning processes. However, bringing a more interventionist approach to learning support within the Finnish educational system would allow more systematic development of instructional practices as well as accumulation, documentation, and distribution of knowledge. Also clear instructions on how to implement these practices are still needed. That is why we have presented a suggestion for providing support in mathematics. However, we are well aware that this suggestion will not be taken seriously as long as the formal documents praise the pedagogical freedom of teachers and local solutions

(as suggested in problem-solving RTI models) for learning difficulties.

A debate about the aims, justification, and uses of the framework of the U.S. RTI (e.g., Artiles et al., 2010; Fuchs et al., 2010; Vaughn et al., 2010; Fuchs and Fuchs, 2017) is still ongoing. Perhaps one way to further clarify the uses of RTI in the US would be that, because they are originally based on the traditions of dynamic pairing of assessment and instruction, they should be seen as a series of carefully selected protocols in the future (Fuchs et al., 2010). This would ensure instructional replicability and flexibility, and the process of identifying learning difficulties would be made clearer.

RECOMMENDATIONS

Within education systems there are always possibilities for improvement, even after reforms take effect. The present analysis contributes to this goal and to the research literature by identifying similarities and differences between two countries with significant experience of RTI-like frameworks. Because formal identification of learning disabilities is not a central part of the current Finnish framework, it is understandable that it resembles those RTI systems that take a problem-solving and consultation-based approach (Ikeda and Gustafson, 2002). A much-welcomed addition to the Finnish RTI would be the data-informed, decision-making and systemized use of standardized assessment and instruction tools, based on systematized progress monitoring (see Fuchs and Fuchs, 2005). This is a question of the allocation of funds that have not been directed toward developing assessment tools and intervention programs in Finland. This is a major difference between Finland and the U.S., where major technical assistance centers, with federal funding, are available to support RTI implementation (e.g., the National Center for Response to Intervention (2016; <http://www.rti4success.org/>); the National Center for Intensive Intervention (National Center for Intensive Intervention; <http://www.intensiveintervention.org>).

If Finland would like to move toward evidence-based or research-based instruction in schools, one of the existing stakeholders (e.g., the Finnish National Board of Education or the Ministry of Culture and Education) should take steps toward establishing similar centers. However, we continuously seek funding to make the www.lukimat.fi service a national RTI center that would be strongly connected to the best universities with the aim of developing evidence-based intervention and advising teachers in addressing learning difficulties.

REFERENCES

- Ahtiainen, R., Beirad, M., Hautamäki, J., Hilasvuori, T., Lintuvuori, M., Thuneberg, H., et al., (2012). *Tehostettua ja Erityistä Tukea Tarvitsevien Oppilaiden Opetuksen Kehittäminen 2007 - 2011. Opetus- ja Kulttuuriministeriön julkaisuja* 2012: 5. [Developing Instruction for Students in Need for Intensified and Special support].
- Artiles, A. J., Bal, A., and King Thorius, K. A. (2010). Back to the future: a critique of response to intervention's social justice

Although the RTI framework seems to be clear, the IDEA legislation leaves too much room for multifaceted interpretations, a situation that leads to, for example, seven-tiered RTI models and the impossibility of comparing the uses of RTI across the U.S. On the other hand, the three-tiered Finnish framework is clear in its background philosophy and purpose (Sabel et al., 2011), but it lacks content: no assessment or intervention tools have been indicated although there are a few available. This lack of indication of materials has led to multiple interpretations of what qualifies as assessment tools (and to discussions if there is a need for using assessment tools at all) and of what intensified or special instruction means.

CONCLUSIONS

What follows from revealing these differing profiles of assessment and instruction within the two countries are some modest suggestions for concluding remarks. For the RTI model used in the U.S., it would be useful to simplify the RTI models in use (see also, Fuchs and Fuchs, 2017) and return to its origins: a three-tiered model with research-based instruction on the first tier, standard protocols on the second tier, and intensive, method-rich, research-based teaching on the third tier. With regards to the future of the Finnish model, the priority, of course, is to collect and create a national resource for assessment materials as well as intervention materials suitable for instructional packages with different intensities and lengths. This process would lead to the use of similar assessment methods and intensified instruction across schools and municipalities and also cumulative knowledge on “what works with whom.” Because the current legislative framework in Finland clearly indicates that support for learning with increasing intensity is required by law, now is a good time to start developing actual assessment policies and ways to implement evidence-based instruction practices intended for the support of learning.

AUTHOR CONTRIBUTIONS

PB has been the major contributor of this paper. LF and DF have provided information on the U.S. RTI. TK and MA have reviewed the paper several times, commented and contributed to the discussion.

FUNDING

This research was funded by ASLA Finnish Fulbright Association g-00009.

views. *Theor. Into Pract.* 49, 250–257. doi: 10.1080/00405841.2010.510447

Basic Education Act (2010). *Basic Education Act: Perusopetuslaki 2010*. Available online at: <http://www.finlex.fi/fi/laki/alkup/2010/20100642> (Finnish, Basic Education Act).

Bender, W., and Shores, C. (2007). *Response to Intervention: A Practical Guide for Every Teacher*. Thousand Oaks, CA: Corwin Press.

Berkeley, S., Bender, W. N., Peaster, L., and Saunders, L. (2009). Implementation of Response to Intervention. A Snapshot of

- progress. *J. Learn. Disabil.* 42, 85–95. doi: 10.1177/0022219408326214
- Björn, P. M., Aro, M., and Koponen, T. (2015). Interventioavastemallien tarjoamat mahdollisuudet kolmiportaisen tuen järjestämiseen: esimerkkinä matematiikan oppimisen tuki [Possibilities of providing support according to response-to-intervention frameworks: supporting mathematics learning]. *NMI Bull.* 25, 3, 10–21.
- Björn, P. M., Aro, M., Koponen, T., Fuchs, L. S., and Fuchs, D. (2016). The many faces of special education within RTI frameworks in the United States and Finland. *Learn. Disabil. Q.* 39, 58–66. doi: 10.1177/0731948715594787
- Bryant, B. R., Bryant, D. P., Porterfield, J., Falcomata, T., Shih, M., Valentine, C., et al. (2014). The effects of a tier 3 intervention for second grade students with serious mathematics difficulties. *J. Learn. Disabil.* 49, 1–13.
- Compton, D. L., Gilbert, J. K., Jenkins, J. R., Fuchs, D., Fuchs, L. S., Cho, E., et al. (2012). Accelerating chronically unresponsive children to Tier 3 instruction: what level of data is necessary to ensure selection accuracy? *J. Learn. Disabil.* 45, 204–216. doi: 10.1177/002221941242151
- Eccles, J. S. (2005). Studying the development of learning and task motivation. *Learn. Instruct.* 15, 161–171. doi: 10.1016/j.learninstruc.2005.04.012
- Elliott, J. G. (2000). “Dynamic assessment in educational contexts: Purpose and promise,” in *Dynamic Assessment: Prevailing Models and Applications*, eds C. Lidz and C. J. G. Elliott (New York, NY: Elsevier), 713–740.
- Elliott, J. G. (2003). Dynamic assessment in educational settings: realising potential. *Edu. Rev.* 55, 15–32. doi: 10.1080/00131910303253
- Fletcher, J. M., and Vaughn, S. (2009). Response to intervention: preventing and remediating academic difficulties. *Child Dev. Perspect.* 3, 30–37. doi: 10.1111/j.1750-8606.2008.00072.x
- Fletcher, J. M., Lyon, G. R., Fuchs, L. S., and Barnes, M. A. (2007). *Learning Disabilities*. New York, NY: Guilford.
- Fuchs, D., and Fuchs, L. S. (2005). Responsiveness-to-intervention: a blueprint for practitioners, policymakers and parents. *Teach. Except. Child.* 38, 57–61. doi: 10.1177/004005990503800112
- Fuchs, D. and Fuchs, L. S. (2006). Introduction to response to intervention: what, why and how valid is it? *Read. Res. Q.* 41, 92–99. doi: 10.1598/RRQ.41.1.4
- Fuchs, D., and Fuchs, L. S. (2017). Critique of the national evaluation of the response to intervention: a case for simpler frameworks. *Except. Child.* 83, 255–268. doi: 10.1177/0014402917693580
- Fuchs, D., Compton, D. L., Fuchs, L. S., Bouton, B., and Caffrey, E. (2011). The construct and predictive validity of a dynamic assessment of young children learning to read: Implications for RTI frameworks. *J. Learn. Disabil.* 44, 339–347. doi: 10.1177/0022219411407864
- Fuchs, D., Fuchs, L. S., Compton, D. L., Bouton, B., Caffrey, E., and Hill, L. (2007). Dynamic assessment as responsiveness to intervention. A scripted protocol to identify young at-risk readers. *Teach. Except. Child.* 39, 58–63. doi: 10.1177/004005990703900508
- Fuchs, D., Fuchs, L. S., and Compton, D. (2012). Smart RTI: A Next-Generation approach to multilevel prevention. *Exceptional Children* 78, 263–279. doi: 10.1177/001440291207800301
- Fuchs, D., Mock, D., Morgan, P. L., and Young, C. L. (2003). Responsiveness-to-intervention: definitions, evidences, and implications for the learning disabilities construct. *Learn. Disabil. Res. Pract.* 18, 157–171. doi: 10.1111/1540-5826.00072
- Fuchs, D., Fuchs, L. S., and Stecker, P. (2010). The “Blurring” of special education in a new continuum of general education placements and services. *Except. Child.* 76, 301–323. doi: 10.1177/001440291007600304
- Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., et al. (2009). *Assisting Students Struggling With Mathematics: Response to Intervention (RTI) for Elementary and Middle Schools* (NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Siteerattu. Available Online at: <https://ies.ed.gov/ncee/wwc/practiceguide/2>
- Grigorenko, E. L. (2009). Dynamic assessment and response to intervention. Two sides of one coin. *J. Learn. Disabil.* 42, 111–132. doi: 10.1177/0022219408326207
- Haager, D., Klingner, J., and Vaughn, S. (eds.). (2007). *Evidence-Based Reading Practices for Response to Intervention*. Baltimore, MD: Paul Brooks Publishing Co.
- IDEA (2004). Individuals with disabilities Education improvement Act. Pub. L. 108–466.
- Ikeda, M., and Gustafson, J. K. (2002). *Heartland AEA 11's Problem Solving Process: Impact on Issues Related to Special Education (Research Rep. No 2002–01)*. Johnston, IA: Heartland Area Education Agency 11.
- Itkonen, T., and Jahnukainen, M. (2007). An analysis of accountability policies in Finland and the United States. *Int. J. Disabil. Dev. Edu.* 54, 5–23.
- Jahnukainen, M. (2011). Different strategies, different outcomes? The history and trends of the inclusive and special education in Alberta (Canada) and in Finland. *Scand. J. Edu. Res.* 55, 489–502. doi: 10.1080/00313831.2010.537689
- Jenkins, J. R., Schiller, E., Blackorby, J., Thayer, S., and Tilly, W. D. (2013). Responsiveness to intervention in reading: architecture and practices. *Learn. Disab. Q.* 36, 36–46. doi: 10.1177/0731948712464963
- Johnson, D. W., Smith, K., and Johnson, R. T., (2006). *Active Learning: Cooperation in the College Classroom*. Edina, MN: Interaction Book Company.
- Lemons, C. J., Fuchs, D., Gilbert, J. K., and Fuchs, L. S. (2014). Evidence-based practices in a changing world: reconsidering the counterfactual in education research. *Educ. Resear.* 43, 242–252. doi: 10.3102/0013189X14539189
- Linnakylä, P., Välijärvi, J., and Arffman, I. (2011). “Finnish basic education: when equity and excellence meet,” in *Equity and Excellence in Education. Towards Maximal Learning Opportunities for All Students*, eds K. Van den Branden, P. Van Avermaet and M. Van Houtte (New York, NY: Routledge) 190–214.
- Mononen, R. (2014). *Early Mathematics Interventions: Supporting Young Children With Low Performance in Mathematics*. Research report 362, University of Helsinki, Helsinki.
- National Association of State Directors of Special Education (2005). Available online at: <http://www.nasdse.org/>
- National Center for Intensive Intervention (2016). Available online at: <http://www.intensiveintervention.org/>
- National Center for Response to Intervention (2016). Available online at: <http://www.rti4success.org/>
- Opetusministeriö. Erityisopetuksen strategia. (2007). *Opetusministeriön Työryhmämuistioita ja Selvityksiä 2007:47. [Strategy for Special Education]*. Ava [http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2007/liitteet/tr47.pdf?lang=](http://www.minedu.fi/export/sites/default/OPM/Julkaisut/2007/liitteet/tr47.pdf?lang=Perusopetuslaki) Perusopetuslaki 2010: <http://www.finlex.fi/fi/laki/alkup/2010/20100642> [Finnish Basic Education Act] (Accessed Oct 20, 2013).
- Pesonen, H., Itkonen, T., Jahnukainen, M., Kontu, E., Kokko, T., Ojala, T., et al. (2015). The implementation of new special education legislation in Finland. *Educ Policy* 29, 162–178. doi: 10.1177/0895904814556754
- Sabel, C., Saxenian, A. L., Miettinen, R., Kristensen, P. H., and Hautamäki, J. (2011). *Individualized Service Provision in the New Welfare State. Lessons From Special Education in Finland*. Helsinki: SITRA Studies 62.
- Sahlberg, P. (2010). Rethinking accountability for a knowledge society. *J. Educ. Change* 11, 45–61. doi: 10.1007/s10833-008-9098-2
- Scholvink, M., and Janssen, L. (2014). *Basisboek RTI. Groei Meten Voor Passend Onderwijs [RTI Measuring Learning Growth in Inclusive Education]*. Huizen: Uitgeverij Pica.
- Seaton, M., Parker, P., Marsh, H. W., Craven, R. G., and Yeung, A. S. (2013). The reciprocal relations between self-concept, motivation and achievement: juxtaposing academic self-concept and achievement goal orientations for mathematics success. *Educ. Psychol.* 34, 49–72. doi: 10.1080/01443410.2013.825232
- Slavin, R. E., and Lake, C. (2008). Effective programs in elementary mathematics: a best-evidence synthesis. *Rev. Educ. Res.* 78, 427–515. doi: 10.3102/0034654308317473
- Vaughn, S., and Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: the promise and potential problems. *Learn. Disabil. Res. Pract.* 18, 137–146. doi: 10.1111/1540-5826.00070
- Vaughn, S., Denton, C. A., and Fletcher, J. M. (2010). Why intensive interventions are necessary for students with severe reading difficulties. *Psychol. Schools* 47, 432–444. doi: 10.1002/pits.20481

- Ysseldyke, J. E., Algozzine, B., and Epps, S. (1983). A logical and empirical analysis of current practices in classifying students as learning disabled. *Except. Children* 50, 160–166. doi: 10.1177/001440298305000207
- Zirkel, P. A. (2012b). The Legal Dimension of RTI: Part I – The Basic Building Blocks. *RTI Action Network*. Available online at: www.rtinetwork.org
- Zirkel, P. A. (2012c). The Legal Dimension of RTI: Part II – State Laws and Guidelines. *RTI Action Network*. Available online at: www.rtinetwork.org
- Zirkel, P. A. (2012a). The legal issues of identification and intervention for K–12 students with dyslexia. *Perspect. Lang. Liter.* 38, 13–16.
- Zirkel, P., and Thomas, L. (2010). State laws for RTI: an updated snapshot. *Teach. Except. Child.* 42, 56–63. doi: 10.1177/004005991004200306

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

Copyright © 2018 Björn, Aro, Koponen, Fuchs and Fuchs. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Do Chinese Children With Math Difficulties Have a Deficit in Executive Functioning?

Xiaochen Wang^{1*}, George K. Georgiou², Qing Li³ and Athanasios Tavouktsoglou⁴

¹ School of Business Administration, Zhejiang Gongshang University, Hangzhou, China, ² Department of Educational Psychology, University of Alberta, Edmonton, AB, Canada, ³ Mental Health Center, Department of Social Sciences and Humanities, Zhejiang University of Media and Communication, Hangzhou, China, ⁴ Faculty of Science, Concordia University of Edmonton, Edmonton, AB, Canada

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

Annemie Desoete,
Ghent University, Belgium
Małgorzata Lipowska,
University of Gdańsk, Poland
Yoshifumi Ikeda,
Joetsu University of Education, Japan

*Correspondence:

Xiaochen Wang
leo197837@163.com

Specialty section:

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

Received: 05 February 2018

Accepted: 18 May 2018

Published: 06 June 2018

Citation:

Wang X, Georgiou GK, Li Q and
Tavouktsoglou A (2018) Do Chinese
Children With Math Difficulties Have
a Deficit in Executive Functioning?
Front. Psychol. 9:906.
doi: 10.3389/fpsyg.2018.00906

Several studies have shown that Executive Functioning (EF) is a unique predictor of mathematics performance. However, whether or not children with mathematics difficulties (MD) experience deficits in EF remains unclear. Thus, the purpose of this study was to examine if Chinese children with MD experience deficits in EF. We assessed 23 children with MD (9 girls, mean age = 10.40 years), 30 children with reading difficulties and MD (RDMD; 12 girls, mean age = 10.82 years), and 31 typically-developing (TD) peers (16 girls, mean age = 10.41 years) on measures of inhibition (Color-Word Stroop, Inhibition), shifting of attention (Planned Connections, Rapid Alternating Stimuli), working memory (Digit Span Backwards, Listening Span), processing speed (Visual Matching, Planned Search), reading (Character Recognition, Sentence Verification), and mathematics (Addition and Subtraction Fluency, Math Standard Achievement Test). The results of MANOVA analyses showed first that the performance of the MD children in all EF tasks was worse than their TD peers. Second, with the exception of the shifting tasks in which the MD children performed better than the RDMD children, the performance of the two groups was similar in all measures of working memory and inhibition. Finally, covarying for the effects of processing speed eliminated almost all differences between the TD and MD groups (the only exception was Listening Span) as well as the differences between the MD and RDMD groups in shifting of attention. Taken together, our findings suggest that although Chinese children with MD (with or without comorbid reading difficulties) experience significant deficits in all EF skills, most of their deficits can be accounted by lower-level deficits in processing speed.

Keywords: executive functioning, math disabilities, working memory, speed of processing, Chinese

INTRODUCTION

Several studies have reported that approximately 20% of school-age children experience mathematics difficulties (MD; see Gross-Tsur et al., 1996; Landerl and Moll, 2010; Geary, 2011; Moll et al., 2014). To better understand the cognitive underpinnings of MD, researchers have further examined the role of several candidate cognitive processes such as general cognitive ability (e.g., Toffalini et al., 2017), working memory (e.g., Passolunghi and Siegel, 2001; Swanson and Kim, 2007), speed of processing (e.g., Koontz and Berch, 1996; Compton et al., 2012), and phonological processing (e.g., Wise et al., 2008; Mazzocco and Grimm, 2013). One of the skills that remain largely unexplored is that of executive functioning (EF). Thus, the purpose of this study was to examine if children with MD experience deficits in EF.

EF is an umbrella term used to represent processes that allow individuals to respond flexibly to our environment and engage in deliberate, goal-directed behavior (e.g., Chan et al., 2008; Diamond, 2013). The three most studied EF skills¹, particularly in relation to mathematics, are inhibition (the ability to suppress distracting information), shifting of attention (the ability to switch between mental sets, representations, and tasks), and working memory (the ability to store information for a short period of time and manipulate or process it)². To date, several studies with typically developing (TD) children have shown that these three EF skills predict (jointly or independently) mathematics performance across a wide range of ages (e.g., Espy et al., 2004; Blair and Razza, 2007; Clark et al., 2010; Lan et al., 2011; Monette et al., 2011; van der Ven et al., 2012; McClelland et al., 2014; Viterbori et al., 2015; Chung et al., 2016; Purpura et al., 2017; see also Friso-van den Bos et al., 2013; Yeniad et al., 2013, for evidence from meta-analyses). In addition, there is some evidence that poor EF correlates with mathematics learning disabilities (e.g., Toll et al., 2011; Willoughby et al., 2016; Morgan et al., 2017).

There are several reasons why inhibition, shifting of attention, and working memory may relate to mathematics (Swanson and Beebe-Frankenberger, 2004; Censabella and Noël, 2008; Bull and Lee, 2014). First, inhibition may help individuals suppress the retrieval and use of developmental immature strategies, inappropriate number bonds (e.g., answering “18” to $3 + 6 = ?$), or the use of information from a word problem that is irrelevant to the solution. Inhibition may also help working memory because inhibition of irrelevant information prevents working memory from becoming overloaded from this information. In turn, shifting of attention may help individuals alternate successfully between mathematical operations, solution strategies, and notations (e.g., between verbal digits, Arabic numerals, and non-symbolic quantity representations), or between the steps involved in solving a multistep problem. Finally, working memory may provide support for strategies such as verbal counting, the direct retrieval of arithmetic facts, the coordination of multiple steps in complex mathematics problems, and the maintenance of interim calculations during mental arithmetic.

Despite the volume of research examining the role of EF skills in TD children (see e.g., Bull and Lee, 2014; Cragg and Gilmore, 2014, for a review), far less is known about the role of EF skills in MD. In addition, the few studies that compared children with MD to TD children have produced mixed findings. Passolunghi et al. (1999; see also Passolunghi and Siegel, 2001, 2004) have found that children who were poor problem solvers were performing significantly lower than good problem solvers in working memory tasks. In addition, as a group, they were

committing more intrusion errors (i.e., they recalled non-target information more often) in a Listening Span task. Based on these findings, Passolunghi and colleagues concluded that the working memory deficits exhibited by children with poor problem solving skills can be traced to more fundamental deficits in inhibition. Because children with MD could not inhibit irrelevant information, more information (relevant or not) was kept active in working memory, which, in turn, overloaded its capacity. Fuchs and colleagues (e.g., Fuchs et al., 2005, 2006) also reported that children with MD had elevated levels of inattentive behavior (based on teacher ratings of attentional skills) and that inattentive behavior, along with working memory deficits, predicted the emergence of computational and problem-solving mathematical difficulties over the course of Grade 1. In contrast to these findings, some studies have shown that children with MD do not experience deficits in inhibition (e.g., van der Sluis et al., 2004; Censabella and Noël, 2005; de Weerd et al., 2013; McDonald and Berg, 2017). For example, van der Sluis et al. (2004) found that children with MD differed from the control group only on tasks involving both inhibition and shifting, and not on tasks involving only inhibition.

Similar contradictory findings have been reported for working memory, and each one of its components (central executive, phonological loop, and visuo-spatial sketchpad; see Passolunghi, 2006, for a review). For example, whereas some researchers have shown that children with MD experience deficits in central executive (e.g., Geary et al., 2000; Passolunghi and Siegel, 2001; Swanson and Sachse-Lee, 2001; Cai et al., 2013), others did not (e.g., McLean and Hitch, 1999; Schuchardt et al., 2008; McDonald and Berg, 2017). Likewise, whereas some researchers have reported significant deficits among children with MD in phonological loop (e.g., Geary et al., 1991; Swanson and Sachse-Lee, 2001; van der Sluis et al., 2005; Cai et al., 2013) and in visuo-spatial sketchpad (e.g., McLean and Hitch, 1999; D’Amico and Guarnera, 2005; Berg, 2008; Cai et al., 2013; Szűcs et al., 2013), others did not (e.g., Bull et al., 1999; Geary et al., 2000, 2004; Landerl et al., 2004).

Several issues need to be considered when interpreting these conflicting results. The first issue relates to speed of processing and whether its effects are partialled out or not before testing for differences between groups in EF. Some studies have shown that children with MD process information more slowly than TD children (e.g., Bull and Johnston, 1997; Swanson and Sachse-Lee, 2001; Chan and Ho, 2010; Vukovic and Siegel, 2010; however, see also Berg, 2008). Because speed of processing is a strong predictor of mathematics performance (e.g., Bull and Johnston, 1997; Fuchs et al., 2006; Peng et al., 2016; Cui et al., 2017), it is possible that MD persist because of persistent deficits in speed of processing, which hinder automatic fact retrieval from long-term memory.

Relatedly, little attention has been paid on how speed of processing has been operationalized in different studies. Some researchers have used naming speed tasks (i.e., digit naming) as measures of speed of processing (e.g., Geary et al., 2007; Chan and Ho, 2010; Vukovic and Siegel, 2010; Moll et al., 2016). Using a naming speed task as a measure of processing speed is problematic because of evidence showing that measures of naming speed do not load on the same factor as measures of speed

¹ We acknowledge that some researchers use the term in a broader way and include under the umbrella of EF skills such cognitive flexibility, verbal fluency, and planning (e.g., Latzman and Markon, 2010; Testa et al., 2012).

² Notice that inhibition, shifting of attention, and working memory are also broad terms (Nigg, 2000; Friedman and Miyake, 2004; Wager et al., 2006; Baddeley, 2012). For example, working memory consists of four components: central executive, phonological loop, visuo-spatial sketchpad, and episodic buffer (Baddeley, 2012).

of processing (e.g., Visual Matching, Cross-out; see van den Bos et al., 2003; Bowey et al., 2004). In addition, poor performance in a naming speed task does not necessarily mean that children experience difficulties in processing speed. Poor performance in a naming speed task may also reflect low-quality phonological representations (e.g., Elbro, 1998), problems in simultaneously processing multiple stimuli when they are presented in serial fashion (e.g., Protopapas et al., 2013), or even problems in forming a “perceptual anchor” in tasks that involve a small set of repeated stimuli (e.g., Ahissar, 2007). Second, because most of the EF tasks (particularly the inhibition and shifting of attention tasks) are speeded and because MD children are often selected based on their poor performance in calculation fluency tasks, this may give rise to EF difficulties. Thus, unless the effects of speed of processing are controlled, we cannot exclude the possibility that the observed difficulties of children with MD in EF tasks are due to lower-level processing speed deficits (see van der Sluis et al., 2004, for a similar argument). In the present study we used more conventional measures of processing speed and we also partialled out the effects of speed of processing prior to testing for group differences in EF skills.

The second issue relates to comorbidity between reading and mathematics. Math difficulties often co-occur with reading difficulties in children with learning disabilities (Gross-Tsur et al., 1996; Moll et al., 2014). Some researchers have argued that children with MD are cognitively different from children with RDMD (e.g., Geary et al., 2007; Landerl et al., 2009; Compton et al., 2012). Because most previous studies did not distinguish between children with MD and children with RDMD (e.g., Geary et al., 2007; Berg, 2008; Censabella and Noël, 2008; Cai et al., 2013), the discrepant findings may reflect a mixed pattern of EF deficits for children with MD and children with RDMD.

It is also worth noting that most previous studies on EF and MD have been conducted in North America or in Europe (see list of studies in the meta-analyses by Friso-van den Bos et al., 2013 and Yeniad et al., 2013), and we do not know if their findings generalize to an East Asian country (e.g., China). We have several reasons to believe that the findings may be different. First, some cross-cultural studies have shown that Chinese children perform better than North American children not only on mathematics skills (e.g., Zhao et al., 2014; Lonnemann et al., 2016; see also Wang and Lin, 2009), but also on EF skills (e.g., Sabbagh et al., 2006; Lan et al., 2011). However, the superior performance of Chinese children in both skills has not led to stronger effects of EF on mathematics skills. Lan et al. (2011), for example, found that whereas inhibition was a unique predictor of calculation ability among American preschoolers, it was not among Chinese preschoolers. Second, Geary et al. (2000) found that American children with MD (with or without comorbid reading difficulties) committed more counting string errors (e.g., recalling the number following one of the addends in the counting string) than their TD peers. They attributed this to inefficient inhibition of irrelevant associations. However, Chinese children practice simple additions and subtractions from the age of 3 (Cheng et al., 2001) and by the time they go to elementary school, they are expected to retrieve the answer to simple calculation problems from their long-term memory. Consequently, Chinese children

with MD may not experience deficits in inhibition. In line with this hypothesis, Peng et al. (2012) found that performance in a color-word Stroop task (one of the most widely used measures of inhibition) did not differentiate Chinese fifth-graders with MD from their TD peers. Finally, the Chinese linguistic system (e.g., short pronunciation of numbers in Chinese and regular number naming structure; see Ng and Rao, 2010, for details) may increase the working memory capacity and reduce the working memory difficulties in Chinese children with MD. Given that only two studies to date have examined the role of EF skills in MD in China (Chan and Ho, 2010; Peng et al., 2012) and none of them has controlled for the effects of speed of processing, more research is needed on this topic.

MATERIALS AND METHODS

Participants

The participants in this study were 84 Chinese children (37 girls, 47 boys; mean age = 10.56 years, $SD = 1.11$) attending five inner-city public schools in Hangzhou. Because there is no formal diagnosis and coding of children as math or reading disabled in China, to select our participants we used the following two-step approach: First, we administered a calculation fluency task (Addition and Subtraction Fluency from WIAT-II; Wechsler, 2009) and a reading fluency task (Sentence Verification; Lei et al., 2011) to a large group of Grade 4, 5, and 6 children ($n = 1,160$; 570 girls and 590 boys). Both tasks were administered to the whole classroom by our trained graduate students.

Second, based on the performance of the children in these two screening measures, we carefully selected three groups of children according to the following criteria: children in the control group had to score at or above the 35th percentile of their grade level in both arithmetic and reading fluency tasks. Children with MD only had to score below the 20th percentile of their grade level in arithmetic fluency (i.e., $\leq$ raw score of 79 in Grade 4; $\leq$ raw score of 92 in Grade 5; $\leq$ raw score of 93 in Grade 6) and above the 35th percentile of their grade level in reading fluency (i.e., $\geq$ raw score of 59 in Grade 4; $\geq$ raw score of 66 in Grade 5; $\geq$ raw score of 67 in Grade 6). Finally, children with both mathematics and reading difficulties had to score below the 20th percentile of their grade level in both mathematics and reading³.

From this selection procedure, we first identified 33 children with poor reading and poor mathematics performance and 25 children with poor mathematics and good reading performance. Three children from the former and two children from the latter group had a non-verbal IQ lower than 85 (assessed with Raven's Matrices) and were excluded from further testing. Thus, our final sample consisted of 30 children with poor mathematics and poor reading performance (18 boys, 12 girls; 10 from Grade 4,

³The 20th and 35th percentiles are commonly used as cutoff scores to select participants with and without reading/mathematics difficulties (e.g., Badian, 1999; Landerl et al., 2004; Fuchs et al., 2008; Tang, 2012). However, as indicated in Swanson and Jerman's (2006) meta-analysis, measures used to establish math disabilities vary from the 48th percentile to the 8th percentile. Geary (2003) also argued that there are no universally agreed-upon criteria for the diagnosis of math difficulties.

8 from Grade 5, and 12 from Grade 6) and 23 children with poor mathematics and good reading performance (14 boys, 9 girls; 9 from Grade 4, 7 from Grade 5, and 7 from Grade 6). Next, to select the children with no reading or MD, we randomly sampled one twentieth of the children meeting the selection criterion (a score higher than the 35th percentile in both reading and mathematics) in each grade⁴. This resulted in 32 children. A child with a non-verbal IQ score lower than 85 was further eliminated leaving us with a sample of 31 children (15 boys, 16 girls; 12 from Grade 4, 10 from Grade 5, and 9 from Grade 6). Parental permission and ethical approval from the Research Ethics Committee of the Zhejiang Gongshang University was obtained prior to testing. Descriptive statistics on age, non-verbal IQ, reading fluency, and arithmetic fluency tasks are presented in Table 1.

Materials

Non-verbal IQ

Non-verbal IQ was assessed with Raven's Standard Progressive Matrices (Raven et al., 2016). Children were presented with a pattern of shapes/geometric designs that was missing a piece and were asked to choose among six to eight alternatives the piece that would accurately complete the pattern. The task consisted of five sets of 12 items (total of 60). A child's score was the total number correct. Cronbach's alpha reliability coefficient in our sample was 0.89.

Speed of Processing

To assess speed of processing we administered two measures: Visual Matching and Planned Search. In Visual Matching, children were required to find and underline the two numbers that were the same in each of the eight rows in a card. There were six numbers of the same length in each row (e.g., 6, 2, 9, 6, 7, 1). In Card 1, the first four rows contained 2-digit numbers and the last four 3-digit numbers. In Card 2, the first four rows contained 4-digit numbers and the last four 5-digit numbers. Maximum time allowed per card was 180 s. A participant's score in Cards 1 and 2 was the total time divided by the number of correct responses. Cronbach's alpha reliability coefficient in our sample was 0.85. In Planned Search, children were asked to

⁴The decision to sample one twentieth of these children was made so that we would have at least as many children in this group as in our second largest group (the group with poor reading and mathematics performance).

match as quickly as possible a target object, number, or letter (located inside a box in the middle of a visual field) with the same object, number, or letter that was located in the visual field among distractors. Each item consisted of two searches, one on the top half of the page and one on the bottom half of the page. Each target had only one match on a page. We recorded the time to complete each search on each page and the participant's score was the total time to complete all searches. Cronbach's alpha reliability coefficient in our sample was 0.79.

Executive Functioning

Inhibition

Inhibition was assessed with the Expressive Attention task from DN CAS-2 (Naglieri et al., 2014) and the Inhibition task from NEPSY-II (Korkman et al., 2007). In Expressive Attention, children were presented with three pages of stimuli. In the first page, children were asked to read color words [i.e., 蓝 (blue), 黄 (yellow), 红 (red), 绿 (green)] that were semi-randomly arranged in eight rows of five. In the second page, children were asked to name an array of color patches of the aforementioned colors. In the third page, children were asked to name as fast as and as accurately as possible the color of the ink in which color words were printed [e.g., the word 红 (Red) printed in blue ink] instead of saying the color word. Before each timed trial, the children were presented with a practice page to ensure they understood the instructions. A participant's score was the time to finish the third page. Cronbach's alpha reliability coefficient in our sample was 0.77. In the Inhibition task, children were required to look at a series of black and white shapes or arrows and name the shape (e.g., say square or circle), the direction (e.g., say up or down), or the opposite (e.g., when you see a square shape, say circle; and when you see a circle shape, say square), depending on the color of the shape or arrow. The completion time in seconds for the test items in each condition (i.e., Naming, Inhibition, and Switching) was recorded. A participant's score was the total time to finish the Inhibition task. Cronbach's alpha reliability coefficient in our sample was 0.82.

Shifting

Shifting was assessed with the Planned Connections task from the DN CAS-2 battery (Naglieri et al., 2014) and the Rapid Alternating Stimuli (RAS) task from the RAN/RAS test battery (Wolf and Denckla, 2005). Planned Connections is a transparent

TABLE 1 | Descriptive statistics on the screening measures separately for each group.

	TD (n = 31)	MD (n = 23)	RDMD (n = 30)		
	M (SD)	M (SD)	M (SD)	F	Group comparisons
Age (years)	10.50 (0.86)	10.40 (1.24)	10.82 (1.11)	1.05	TD = RDMD = MD
Raven	105.42 (6.35)	102.30 (5.57)	102.38 (5.34)	1.41	TD = RDMD = MD
Sentence Verification	72.29 (4.58)	69.35 (6.54)	38.60 (6.92)	227.44***	TD = MD > RDMD
Character Recognition	142.65 (3.77)	139.43 (8.48)	124.50 (9.06)	51.34***	TD = MD > RDMD
Calculation Fluency	135.45 (17.32)	80.00 (10.02)	70.93 (17.11)	115.50***	TD > MD = RDMD
MSAT	35.61 (7.49)	32.70 (7.25)	29.77 (5.86)	5.51***	TD > MD = RDMD

TD, typically developing group; MD, mathematics difficulties group; RDMD, reading and mathematics difficulties group. *** $p < 0.001$.

adaptation of Trail Making (Reitan and Wolfson, 1992). In this task, children were presented with four pages of numbers (1–13) and letters (A–M), and, in each page, they were asked to connect the numbers to the letters in successive order (1, A, 2, B, 3, C, etc.) as fast and as accurately as possible. The score was the total time to finish all pages. Cronbach's alpha reliability in our study was 0.80. In RAS, children were required to name as fast and as accurately as possible color patches mixed up with digits (i.e., blue, 2, yellow, 6, green, 9, black, 7, etc.). The color patches and digits were randomly presented in five rows of ten. Prior to testing, the children were asked to name each of the RAS stimuli in a practice trial to ensure familiarity. A participant's score was the total time to name all items. Wolf and Denckla (2005) reported test-retest reliability for this task to be 0.84.

Working memory

The Digit Span Backwards task and the Listening Span task were used to assess working memory. The Digit Span Backwards task was adopted from WISC-III (Wechsler, 1992). Children were asked to repeat a sequence of digits in the reverse order. The strings of digits were presented orally by the experimenter with a time interval of about 1 second between each digit. The strings started with only two digits and one digit was added at each difficulty level (the maximum length was seven digits). The task was terminated when children failed both trials of a given length. The children's score was the number of digit strings accurately recalled. Cronbach's alpha reliability coefficient in our sample was 0.75. The Listening Span task was adapted in Chinese from Daneman and Carpenter (1980). The children listened to groups of sentences and were asked to determine if each sentence was true or false (e.g., "All mothers work in an office"). Children were instructed to keep the last word in each sentence in their memory and then after completing a sentence group they were asked to say the last word in each sentence in the same order. A participant's score was the total number of sets correctly recalled (max = 5). Cronbach's alpha reliability coefficient in our sample was 0.81.

Arithmetic Skills

Arithmetic accuracy

The Math Standard Achievement Test (MSAT) from Dong and Lin (2011) was used to assess arithmetic accuracy. The task has been used in previous studies in China showing good psychometric properties (e.g., Cai et al., 2013, 2018). The test included 30 items: 26 items were multiple choice questions (e.g., If $\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$ is number 31, what number is $\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$? 4, 9, 45, or 54?) and 4 items were fill-in questions (e.g., Based on the map you have in front of you, how long will it take Fang to go to the bookstore, if he first passes by Li's home?). The task was discontinued after four consecutive errors. A participant's score was the total number correct. Cronbach's alpha reliability coefficient in our sample was 0.84.

Arithmetic fluency

To assess arithmetic fluency we administered the addition and subtraction fluency tasks from WIAT-III (Wechsler, 2009). In each subtest, children were asked to solve as many additions or subtractions as possible within a 1-min time limit by writing their

response in the space provided beside each problem. Each subtest included two pages (24 items on each page; total of 48 problems). A participant's score was the total correct number of additions and subtractions completed within the time limit. The scores in addition fluency correlated 0.85 with the scores in subtraction fluency.

Reading Skills

Reading accuracy

Character Recognition was adopted from Li et al. (2012) to assess reading accuracy. The task has been used in previous studies in Chinese showing good reliability and validity evidence (e.g., Xue et al., 2013; Zhang et al., 2013; Liao et al., 2015). Children were asked to read aloud 150 Chinese two-character words that were arranged in terms of increasing difficulty. The task was discontinued after 15 consecutive errors. A participant's score was the total number correct. Cronbach's alpha reliability coefficient in our sample was 0.89.

Reading fluency

Sentence Verification from Lei et al. (2011) was used to assess reading fluency. The task has been used in previous studies in Chinese showing good psychometric properties (e.g., Liao et al., 2014; Pan et al., 2016; Xia et al., 2018). Children were asked to read silently simple sentences and indicate if the meaning of each sentence was true or false by circling Y (for Yes) or N (for No) printed at the end of each sentence (e.g., *Horse is an animal.* Y – N). The semantic content and linguistic format of each sentence was simple so that only very basic comprehension processes were required. A 3-min time limit was implemented. A participant's score was the total number of correctly answered sentences.

Procedure

Testing was conducted by the first and third authors, and six graduate students who received extensive training on test administration and scoring. Sentence Verification, and Addition and Subtraction Fluency were administered in a group setting. The rest of the tasks were administered to each child individually during school hours in a quiet room at school. Individual testing took place 3 weeks following the group testing and lasted approximately an hour.

RESULTS

Group Comparisons on Inhibition

First, we ran a MANOVA with the two inhibition tasks as dependent variables and group as a fixed factor. The results revealed a main effect of group, Wilk's $\lambda = 0.610$, $F(4,160) = 11.23$, $p < 0.001$. Follow-up ANOVAs showed that the groups differed in both Expressive Attention [$F(2,81) = 18.85$, $p < 0.001$] and Inhibition [$F(2,81) = 18.90$, $p < 0.001$]. *Post hoc* analyses showed that the TD group performed better than the MD and RDMD groups in both Expressive Attention and Inhibition (see **Table 2**). No significant differences were found between the MD and the RDMD groups.

TABLE 2 | Descriptive statistics on the EF tasks and group comparisons.

	TD (<i>n</i> = 31)			MD (<i>n</i> = 23)			RDMD (<i>n</i> = 30)			Processing speed as a covariate		
	<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>		<i>M</i>	<i>SD</i>		<i>F</i> (2,81)	η^2	<i>Post hoc</i>
Inhibition												
Exp. Attention	49.88	9.44		63.56	15.83		72.13	16.91				
Inhibition	30.16	4.15		36.84	6.93		41.27	9.31		4.62*	0.10	TD = MD < RDMD
Shifting												
Pl. Connections	37.57	6.36		51.02	18.24		60.36	14.89		3.99*	0.09	TD = MD < RDMD
RAS	29.90	4.87		37.62	9.86		44.45	12.29				
Working memory												
DSB	8.16	2.56		6.39	1.95		5.67	1.92		4.83*	0.11	TD = MD, RDMD = MD, TD < RDMD
Listening Span	3.03	0.84		2.43	0.90		2.23	1.16		4.84*	0.11	TD = MD, RDMD = MD, TD < RDMD
Processing speed												
Visual Matching	48.58	4.50		42.83	5.254		39.83	5.849		3.12*	0.07	TD = MD, RDMD = MD, TD > RDMD
Planned Search	7.20	1.31		8.49	2.65		8.60	1.70		5.60*	0.15	TD > MD = RDMD

TD, typically developing group; MD, mathematics difficulties group; RDMD, reading and mathematics difficulties group; DBS, Digit Span Backwards; RAS, Rapid Alternating Stimuli. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Group Comparisons on Shifting

A MANOVA with the two shifting tasks as dependent variables and group as a fixed factor revealed a main effect of group, Wilk's $\lambda = 0.579$, $F(4,160) = 12.57$, $p < 0.001$. Follow-up ANOVAs showed that the groups differed in both Planned Connections [$F(2,81) = 21.65$, $p < 0.001$] and RAS [$F(2,81) = 17.80$, $p < 0.001$]. *Post hoc* analyses showed that the TD group performed better than the MD and RDMD groups in both Planned Connections and RAS. The MD group also performed significantly better than the RDMD group (see Table 2).

Group Comparisons on Working Memory

Next, we ran a MANOVA with the two working memory tasks as dependent variables and group as a fixed factor. The results revealed a main effect of group, Wilk's $\lambda = 0.745$, $F(4,160) = 6.34$, $p < 0.01$. Follow-up ANOVAs showed that the groups differed in both Digit Span Backwards [$F(2,81) = 10.46$, $p < 0.001$] and Listening Span [$F(2,81) = 8.04$, $p < 0.01$]. *Post hoc* analyses showed that the TD group obtained significantly higher scores than the MD and RDMD groups in both Digit Span Backwards and Listening Span. There was no significant difference between the MD and RDMD groups (see Table 2).

Group Comparisons on Processing Speed

Finally, we ran a MANOVA with the two processing speed tasks as dependent variables and group as a fixed factor. The results revealed a main effect of group, Wilk's $\lambda = 0.640$, $F(4,160) = 10.01$, $p < 0.001$. Follow-up ANOVAs showed that the groups differed in both Visual Matching [$F(2,81) = 22.03$, $p < 0.001$] and Planned Search [$F(2,81) = 4.99$, $p < 0.05$]. *Post hoc* analyses showed that the TD group obtained significantly higher scores than the MD and RDMD groups in both Visual Matching and Planned Search. Again, there was no significant difference between the MD and RDMD groups (see Table 2).

Group Comparisons on Executive Functioning After Controlling for Processing Speed

Finally, we performed three MANCOVAs (one for each EF skill) covarying for the effects of processing speed (Visual Matching and Planned Search) (see Table 2). In terms of inhibition, the results revealed a main effect of group, Wilk's $\lambda = 0.861$, $F(4,158) = 3.07$, $p < 0.001$. Follow-up ANCOVAs showed that the groups differed in both Expressive Attention [$F(2,80) = 4.62$, $p < 0.05$] and Inhibition [$F(2,80) = 3.99$, $p < 0.05$]. *Post hoc* analyses revealed no significant differences between the TD and MD groups. In addition, both groups performed significantly better than the RDMD group.

In terms of shifting, the results of MANCOVA revealed a main effect of group, Wilk's $\lambda = 0.832$, $F(4,158) = 3.80$, $p < 0.001$. Follow-up ANCOVAs showed that the groups differed in both Planned Connections [$F(2,80) = 4.83$, $p < 0.05$] and RAS [$F(2,80) = 4.84$, $p < 0.05$]. *Post hoc* analyses revealed that the only significant difference was between the TD and the RDMD

groups. The MD group did not differ significantly from the TD and RDMD groups.

Finally, in terms of working memory, the results of MANCOVA revealed a main effect of group, Wilk's $\lambda = 0.858$, [$F(4,158) = 3.14$, $p < 0.001$]. Follow-up ANCOVAs showed that the groups differed in both Digit Span Backwards [$F(2,80) = 3.12$, $p < 0.05$] and Listening Span, [$F(2,80) = 5.60$, $p < 0.05$]. *Post hoc* analyses showed no significant differences between the TD and MD groups in Digit Span Backwards, but significant difference between the two groups in Listening Span. The MD group did not differ from the RDMD group on either working memory task.

DISCUSSION

Several studies have demonstrated that EF is an important predictor of mathematics performance and a risk factor of MD (see Bull and Lee, 2014; Cragg and Gilmore, 2014, for reviews). Nevertheless, because EF consists of several sub-components (inhibition, shifting, and working memory; Miyake et al., 2000) and because MD overlaps with reading difficulties (Landerl and Moll, 2010), it remains unclear if MD children have a deficit in all EF sub-components and if these deficits are accentuated by concomitant difficulties in reading disabilities. Thus, in this study, we sought to examine the nature of EF deficits in Chinese children with mathematics disabilities (with or without comorbid reading difficulties).

First, our results showed that the MD children differed from their TD peers on all EF skills (see Chan and Ho, 2010; Cai et al., 2013, for similar findings). However, most differences disappeared once we controlled for the effects of speed of processing. Notably, the only difference between the TD and MD groups that remained significant was in Listening Span. This suggests that the significant differences between MD and TD children in inhibition or shifting of attention that have been reported in previous studies (e.g., Geary et al., 2000, 2007; Szűcs et al., 2013; McDonald and Berg, 2017) may reflect differences between groups in speed of processing more so than in EF. The argument put forward by some researchers that deficits in inhibition are likely responsible for the observed deficits of MD children in working memory (e.g., Passolunghi et al., 1999; Passolunghi and Siegel, 2001) does not seem to be supported by our findings either, because the MD group continued to perform more poorly than the TD group in Listening Span despite their equal performance in inhibition (that is after controlling for processing speed).

The absence of a significant difference between the TD and MD groups in our study may also reflect cultural differences. More specifically, because Chinese children go to school at the age of 3 and they systematically practice simple additions/subtractions, by the time they go to Grade 1 they are able to retrieve the answer to simple calculation problems from their memory (Geary et al., 1996). This likely reduces interference from competing responses and decreases the effect of inhibition in mathematics. In addition, because Chinese digits have a shorter pronunciation duration than digits in other languages such as English (digits in China are monosyllabic) and because shorter names allow for a greater number of digits to be stored

in working memory, this may explain why we did not observe any differences between the MD and TD groups in Digit Span Backwards.

Second, our findings replicate those of previous studies in North America/Europe (e.g., Swanson, 1993; Geary et al., 2000; Passolunghi and Siegel, 2001; Swanson and Sachse-Lee, 2001; Andersson and Lyxell, 2007) showing persistent deficits of the MD group in working memory (at least on measures of the central executive such as Listening Span). Although similar differences were detected in Digit Span Backwards, they did not survive the statistical control of speed of processing. An explanation may relate to the nature of the Digit Span Backwards task. More specifically, some researchers have argued that although it is frequently used as a measure of working memory, it is relatively shallow in its processing demands (e.g., St Clair-Thompson, 2010; Georgiou and Das, 2016).

Some limitations of the present study are worth mentioning. First, our participants did not come with a formal diagnosis of learning disabilities. Such a diagnosis does not exist in China. That is also why we selected our participants by first screening a relatively large sample of children. Second, we screened our participants using reading and mathematics fluency tasks in a group setting. Despite the fact that this is a convenient approach and it has been used in several previous studies to screen children for learning disabilities, we acknowledge that it comes with limitations (e.g., some children may get distracted in the presence of other students and may not invest the maximum of their effort). Nonetheless, we obtained similar differences between groups in the two individually administered measures (Character Recognition and MSAT), which allows us to say with some confidence that our selection worked quite well. Third, we did not manipulate the modality of the EF tasks (i.e., verbal vs. quantity) in our study. For example, some previous studies have used either verbal or numerical EF tasks to rule out the possibility that EF deficits manifest themselves only within a specific domain (e.g., Peng et al., 2013). Fourth, because of limited resources, we did not assess children with only reading difficulties. We acknowledge that this would have strengthened the findings of our study. Finally, we espoused a rather narrow view of EF consisting of three core components. That is why we did not administer measures of planning or visual-spatial memory and we did not assess other types of inhibition that are often operationalized with Go/No-go tasks. Future studies may assess EF more broadly.

CONCLUSION

To conclude, our findings add to a growing body of research on the role of EF skills in MD (e.g., Swanson, 1993; Passolunghi and Siegel, 2001; van der Sluis et al., 2004; Geary et al., 2007; Compton et al., 2012; Peng et al., 2012) highlighting the role of speed of processing as a mediating factor in the severity of EF difficulties. Importantly, although our MD group differed from the TD group on all EF tasks, after controlling for speed of processing, the only difference that remained significant was in Listening Span. Although we are not the first ones to report non-significant differences between children with and without MD in inhibition and/or shifting of attention (see e.g.,

van der Sluis et al., 2004; Censabella and Noël, 2008 for similar findings), our study is the first one to show this in mainland China. We argue here that the linguistic features of Chinese (i.e., short pronunciation of digits in Chinese, transparent number-naming system), the age at which children start learning to do simple calculations, and the increased levels of parental involvement in children's learning (particularly in mathematics; see Deng et al., 2015) may alleviate the negative impact of EF difficulties in MD in China. Future studies may examine the role of different EF skills in MD across cultures.

AUTHOR CONTRIBUTIONS

XW, GG, and AT designed the study. XW and QL helped in the data collection and data analysis, and also wrote the “Materials

and Methods” and “Results” section of the paper. GG wrote the “Introduction” and “Discussion” sections of the paper. AT helped in revising the whole manuscript.

FUNDING

This study was supported by a grant from Scientific Planning of the Zhejiang Philosophy Society (Project No. 14NDJC110YB).

ACKNOWLEDGMENTS

We would like to thank Gao Xinjie, Yinying, Li Jiahao, and Wu Yinfei at Zhejiang Gongshang University for their assistance in data collection.

REFERENCES

- Ahissar, M. (2007). Dyslexia and the anchoring-deficit hypothesis. *Trends Cogn. Sci.* 11, 458–465. doi: 10.1016/j.tics.2007.08.015
- Andersson, U., and Lyxell, B. (2007). Working memory deficit in children with mathematical difficulties: a general or specific deficit? *J. Exp. Child Psychol.* 96, 197–228. doi: 10.1016/j.jecp.2006.10.001
- Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29. doi: 10.1146/annurev-psych-120710-100422
- Badian, N. (1999). Persistent arithmetic, reading or arithmetic and reading disability. *Ann. Dyslexia* 49, 43–70. doi: 10.1007/s11881-999-0019-8
- Berg, D. H. (2008). Cognitive impairments of children with severe arithmetic difficulties: cognitive deficit or developmental lag? *Except. Educ. Canada* 18, 59–92.
- Blair, C., and Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Dev.* 78, 647–663. doi: 10.1111/j.1467-8624.2007.01019.x
- Bowey, J. A., Storey, T., and Ferguson, A. N. (2004). The association between continuous naming speed and word reading skills in fourth- to sixth-grade children. *Aust. J. Psychol.* 56, 155–163. doi: 10.1080/00049530412331283345
- Bull, R., and Johnston, R. S. (1997). Children's arithmetical difficulties: contributions from processing speed, item identification, and short-term memory. *J. Exp. Child Psychol.* 65, 1–24. doi: 10.1006/jecp.1996.2358
- Bull, R., Johnston, R. S., and Roy, J. A. (1999). Exploring the roles of the visual-spatial sketch pad and central executive in children's arithmetical skills: views from cognition and developmental neuropsychology. *Dev. Neuropsychol.* 15, 421–442. doi: 10.1080/87565649909540759
- Bull, R., and Lee, K. (2014). Executive functioning and mathematics achievement. *Child Dev. Perspect.* 8, 36–41. doi: 10.1111/cdep.12059
- Cai, D., Li, Q. W., and Deng, C. P. (2013). Cognitive processing characteristics of 6th to 8th grade Chinese students with mathematics learning disability: relationships among working memory, pass processes, and processing speed. *Learn. Individ. Dif.* 27, 120–127. doi: 10.1016/j.lindif.2013.07.008
- Cai, D., Viljaranta, J., and Georgiou, G. K. (2018). Direct and indirect effects of self-concept of ability on math skills. *Learn. Individ. Dif.* 61, 51–58. doi: 10.1016/j.lindif.2017.11.009
- Censabella, S., and Noël, M. P. (2005). The inhibition of exogenous distracting information in children with learning disabilities. *J. Learn. Disabil.* 38, 400–410. doi: 10.1177/00222194050380050301
- Censabella, S., and Noël, M. P. (2008). The inhibition capacities of children with mathematical disabilities. *Child Neuropsychol.* 14, 1–20. doi: 10.1080/09297040601052318
- Chan, B. M., and Ho, C. S. (2010). The cognitive profile of Chinese children with mathematics difficulties. *J. Exp. Child Psychol.* 107, 260–279. doi: 10.1016/j.jecp.2010.04.016
- Chan, R. C. K., Shum, D., Touloupoulou, T., and Chen, E. Y. H. (2008). Assessment of executive functions: review of instruments and identification of critical issues. *Arch. Clin. Neuropsychol.* 23, 201–216. doi: 10.1016/j.acn.2007.08.010
- Cheng, Z. J., Chan, K. S., Li, Y. L., Ng, S. N., and Woo, Y. S. (2001). Preschool children's actual computational ability. *Educ. J.* 29, 121–135.
- Chung, K. K. H., Liu, H., McBride, C., Wong, A. M. Y., and Lo, J. C. M. (2016). How socioeconomic status, executive functioning and verbal interactions contribute to early academic achievement in Chinese children. *Educ. Psychol.* 37, 1–19.
- Clark, C. A. C., Pritchard, V. E., and Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Dev. Psychol.* 46, 1176–1191. doi: 10.1037/a0019672
- Compton, D. L., Fuchs, L. S., Fuchs, D., Lambert, W., and Hamlett, C. (2012). The cognitive and academic profiles of reading and mathematics learning disabilities. *J. Learn. Disabil.* 45, 79–95. doi: 10.1177/0022219410393012
- Cragg, L., and Gilmore, C. (2014). Skills underlying mathematics: the role of executive function in the development of mathematics proficiency. *Trends Neurosci. Educ.* 3, 63–68. doi: 10.1016/j.tine.2013.12.001
- Cui, J., Georgiou, G. K., Zhang, Y., Li, Y., Shu, H., and Zhou, X. (2017). Examining the relationship between rapid automatized naming and arithmetic fluency in Chinese kindergarten children. *J. Exp. Child Psychol.* 154, 146–163. doi: 10.1016/j.jecp.2016.10.008
- D'Amico, A., and Guarnera, M. (2005). Exploring working memory in children with low arithmetical achievement. *Learn. Individ. Dif.* 15, 189–202. doi: 10.1016/j.lindif.2005.01.002
- Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verbal Learning Verbal Behav.* 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6
- de Weerd, F., Desoete, A., and Roeyers, H. (2013). Behavioral inhibition in children with learning disabilities. *Res. Dev. Disabil.* 34, 1998–2007. doi: 10.1016/j.ridd.2013.02.020
- Deng, C.-P., Silinskas, G., Wei, W., and Georgiou, G. (2015). Cross-lagged relationships between home literacy/numeracy environment and academic achievement in Chinese. *Early Child. Res. Q.* 33, 12–20. doi: 10.1016/j.jecresq.2015.05.001
- Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750
- Dong, D., and Lin, C. (2011). *The Standard Tests for Psychological Development of Chinese Children and Adolescents*. Beijing: Science Press.
- Elbro, C. (1998). When reading is “readn” or somthn: distinctness of phonological representations of lexical items in normal and disabled readers. *Scand. J. Psychol.* 39, 149–153. doi: 10.1111/1467-9450.393070
- Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., and Senn, T. E. (2004). The contribution of executive functions to emergent mathematic

- skills in preschool children. *Dev. Neuropsychol.* 26, 465–486. doi: 10.1207/s15326942dn2601_6
- Friedman, N. P., and Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *J. Exp. Psychol. Gen.* 133, 101–135. doi: 10.1037/0096-3445.133.1.101
- Friso-van den Bos, I., van der Ven, S. H. G., Kroesbergen, E. H., and van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: a meta-analysis. *Educ. Res. Rev.* 10, 29–44. doi: 10.1016/j.edurev.2013.05.003
- Fuchs, L., Compton, D., Fuchs, D., Paulsen, K., Bryant, J. D., and Hamlett, C. L. (2005). The prevention, identification, and cognitive determinants of math difficulty. *J. Educ. Psychol.* 97, 493–513. doi: 10.1037/0022-0663.97.3.493
- Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *J. Educ. Psychol.* 98, 29–43. doi: 10.1037/0022-0663.98.1.29
- Fuchs, L. S., Seethaler, P. M., Powell, S. R., Fuchs, D., Hamlett, C. L., and Fletcher, J. M. (2008). Effects of preventative tutoring on the mathematical problem solving of third-grade students with math and reading difficulties. *Except. Child.* 74, 155–173. doi: 10.1177/001440290807400202
- Geary, D. C. (2003). “Learning disabilities in arithmetic,” in *Handbook of Learning Disabilities*, eds H. L. Swanson, K. Harris, and S. Graham (New York, NY: Guilford), 199–212.
- Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Dev. Psychol.* 47, 1539–1552. doi: 10.1037/a0025510
- Geary, D. C., Bow-Thomas, C. C., Liu, F., and Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: influence of age, language, and schooling. *Child. Dev.* 67, 2022–2044. doi: 10.2307/1131607
- Geary, D. C., Brown, S. C., and Samaranayake, V. A. (1991). Cognitive addition: a short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Dev. Psychol.* 27, 787–797. doi: 10.1037/0012-1649.27.5.787
- Geary, D. C., Hamson, C. O., and Hoard, M. K. (2000). Numerical and arithmetical cognition: a longitudinal study of process and concept deficits in children with learning disability. *J. Exp. Child Psychol.* 77, 236–263. doi: 10.1006/jecp.2000.2561
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., and DeSoto, M. C. (2004). Strategy choices in simple and complex addition: contributions of working memory and counting knowledge for children with mathematical disability. *J. Exp. Child Psychol.* 88, 121–151. doi: 10.1016/j.jecp.2004.03.002
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., and Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Dev.* 78, 1343–1359. doi: 10.1111/j.1467-8624.2007.01069.x
- Georgiou, G., and Das, J. P. (2016). What component of executive functions contributes to normal and impaired reading comprehension in young adults? *Res. Dev. Disabil.* 49–50, 118–128. doi: 10.1016/j.ridd.2015.12.001
- Gross-Tsur, V., Manor, O., and Shalev, R. S. (1996). Developmental dyscalculia: prevalence and demographic features. *Dev. Med. Child Neurol.* 38, 25–33. doi: 10.1111/j.1469-8749.1996.tb15029.x
- Koontz, K. L., and Berch, D. B. (1996). Identifying simple numerical stimuli: processing inefficiencies exhibited by arithmetic learning disabled children. *Math. Cogn.* 2, 1–23. doi: 10.1080/135467996387525
- Korkman, M., Kirk, U., and Kemp, S. L. (2007). *NEPSY II. Administrative Manual*. San Antonio, TX: Psychological Corporation.
- Lan, X., Legare, C. H., Ponitz, C. C., Li, S., and Morrison, F. J. (2011). Investigating the links between the subcomponents of executive function and academic achievement: a cross-cultural analysis of Chinese and American preschoolers. *J. Exp. Child Psychol.* 108, 677–692. doi: 10.1016/j.jecp.2010.11.001
- Landerl, K., Bevan, A., and Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8–9 year old students. *Cognition* 93, 99–125. doi: 10.1016/j.cognition.2003.11.004
- Landerl, K., Fussenegger, B., Moll, K., and Willburger, E. (2009). Dyslexia and dyscalculia: two learning disorders with different cognitive profiles. *J. Exp. Child Psychol.* 103, 309–324. doi: 10.1016/j.jecp.2009.03.006
- Landerl, K., and Moll, K. (2010). Comorbidity of learning disorders: Prevalence and familial transmission. *J. Child Psychol. Psychiatry* 51, 287–294. doi: 10.1111/j.1469-7610.2009.02164.x
- Latzman, R. D., and Markon, K. E. (2010). The factor structure and age-related factorial invariance of the Delis-Kaplan executive function system (D-KEFS). *Assessment* 17, 172–184. doi: 10.1177/1073191109356254
- Lei, L., Pan, J., Liu, H., McBride-Chang, C., Li, H., Zhang, Y., et al. (2011). Developmental trajectories of reading development and impairment from ages 3 to 8 years in Chinese children. *J. Child Psychol. Psychiatry* 52, 212–220. doi: 10.1111/j.1469-7610.2010.02311.x
- Li, H., Shu, H., McBride-Chang, C., Liu, H., and Peng, H. (2012). Chinese children’s character recognition: visuo-orthographic, phonological processing and morphological skills. *J. Res. Read.* 35, 287–307. doi: 10.1111/j.1467-9817.2010.01460.x
- Liao, C.-H., Altani, A., and Georgiou, G. (2014). “Rapid automatized naming and reading in Chinese: a highly specific and stable relationship,” in *Poster presented at the 23rd biennial meeting of the International Society for the Study of Behavioural Development*, Shanghai.
- Liao, C.-H., Deng, C.-P., Hamilton, J., Lee, S.-C., Wei, W., and Georgiou, G. (2015). The role of RAN in reading development and dyslexia in Chinese. *J. Exp. Child Psychol.* 130, 106–122. doi: 10.1016/j.jecp.2014.10.002
- Lonnemann, J., Li, S., Zhao, P., Li, P., Linkersdörfer, J., Lindberg, S., et al. (2016). Differences in arithmetic performance between Chinese and German adults are accompanied by differences in processing of non-symbolic numerical magnitude. *Front. Psychol.* 7:1337. doi: 10.3389/fpsyg.2016.01337
- Mazzocco, M. M., and Grimm, K. J. (2013). Growth in rapid automatized naming from Grades k – 8 in children with math or reading disabilities. *J. Learn. Disabil.* 46, 517–533. doi: 10.1177/0022219413477475
- McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acock, A. C., Miao, A., et al. (2014). Predictors of early growth in academic achievement: the head-toes-knees-shoulders task. *Front. Psychol.* 5:599. doi: 10.3389/fpsyg.2014.00599
- McDonald, P. A., and Berg, D. H. (2017). Identifying the nature of impairments in executive functioning and working memory of children with severe difficulties in arithmetic. *Child Neuropsychol.* 1, 1–16. doi: 10.1080/09297049.2017.1377694
- McLean, J. F., and Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *J. Exp. Child Psychol.* 74, 240–260. doi: 10.1006/jecp.1999.2516
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cognit. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734
- Moll, K., Gabel, S. M., Gooch, D., Landerl, K., and Snowling, M. J. (2016). Cognitive risk factors for specific learning disorder: processing speed, temporal processing, and working memory. *J. Learn. Disabil.* 49, 272–281. doi: 10.1177/0022219414547221
- Moll, K., Kunze, S., Neuhoﬀ, N., Bruder, J., and SchulteKörne, G. (2014). Specific learning disorder: prevalence and gender differences. *PLoS One* 9:e103537. doi: 10.1371/journal.pone.0103537
- Monette, S., Bigras, M., and Guay, M. C. (2011). The role of the executive functions in school achievement at the end of Grade 1. *J. Exp. Child Psychol.* 109, 158–173. doi: 10.1016/j.jecp.2011.01.008
- Morgan, P. L., Li, H., Farkas, G., Cook, M., Pun, W. H., and Hillemeier, M. M. (2017). Executive functioning deficits increase kindergarten children’s risk for reading and mathematics difficulties in first grade. *Contemp. Educ. Psychol.* 50, 23–32. doi: 10.1016/j.cedpsych.2016.01.004
- Naglieri, J. A., Das, J. P., and Goldstein, S. (2014). *Cognitive Assessment System*, 2nd Edn. Rolling Meadows, IL: Riverside Publishing.
- Ng, S. S. N., and Rao, N. (2010). Chinese number words, culture, and mathematics learning. *Rev. Educ. Res.* 80, 180–206. doi: 10.3102/0034654310364764

- Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: views from cognitive and personality psychology and a working inhibition taxonomy. *Psychol. Bull.* 126, 200–246. doi: 10.1037/0033-2909.126.2.220
- Pan, J., Song, S., Su, M. M., McBride, C., Liu, H. Y., Zhang, Y. P., et al. (2016). On the relationship between phonological awareness, morphological awareness and Chinese literacy skills: evidence from an 8-year longitudinal study. *Dev. Sci.* 19, 982–991. doi: 10.1111/desc.12356
- Passolunghi, M. C. (2006). “Working memory and arithmetic learning disability,” in *Working Memory and Neurodevelopmental Disorders*, ed. T. P. Alloway (Hove: Psychology Press), 113–138.
- Passolunghi, M. C., Cornoldi, C., and De Liberto, S. (1999). Working memory and intrusions of irrelevant information in a group of specific poor problem solvers. *Mem. Cognit.* 27, 779–790. doi: 10.3758/BF03198531
- Passolunghi, M. C., and Siegel, L. S. (2001). Short-term memory, working memory, and inhibitory control in children with difficulties in arithmetic problem solving. *J. Exp. Child Psychol.* 80, 44–57. doi: 10.1006/jecp.2000.2626
- Passolunghi, M. C., and Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *J. Exp. Child Psychol.* 88, 348–367. doi: 10.1016/j.jecp.2004.04.002
- Peng, P., Namkung, J. M., Fuchs, D., Fuchs, L. S., Patton, S., and Yen, L. (2016). A longitudinal study of predictors of early calculation development among young children at risk for learning difficulties. *J. Exp. Child Psychol.* 152, 221–241. doi: 10.1016/j.jecp.2016.07.017
- Peng, P., Sha, T., and Beilei, L. (2013). The deficit profile of working memory, inhibition and updating in Chinese children with reading difficulties. *Learn. Individ. Dif.* 25, 111–117. doi: 10.1016/j.lindif.2013.01.012
- Peng, P., Sun, C., Li, B., and Sha, T. (2012). Phonological storage and executive function deficits in children with mathematics difficulties. *J. Exp. Child Psychol.* 112, 452–466. doi: 10.1016/j.jecp.2012.04.004
- Protopapas, A., Altani, A., and Georgiou, G. K. (2013). Development of serial processing in reading and rapid naming. *J. Exp. Child Psychol.* 116, 914–926. doi: 10.1016/j.jecp.2013.08.004
- Purpura, D. J., Schmitt, S. A., and Ganley, C. M. (2017). Foundations of mathematics and literacy: the role of executive functioning components. *J. Exp. Child Psychol.* 153, 15–34. doi: 10.1016/j.jecp.2016.08.010
- Raven, J. C., Court, J. H., and Raven, J. C. (2016). *Manual for Raven's Progressive Matrices and Vocabulary Scales, Section 3 Standard Progressive Matrices*. San Antonio, TX: Harcourt Assessment.
- Reitan, R. M., and Wolfson, D. (1992). Conventional intelligence measurements and neuropsychological concepts of adaptive abilities. *J. Clin. Psychol.* 48, 521–529. doi: 10.1002/1097-4679(199207)48:4<521::AID-JCLP2270480414>3.0.CO;2-C
- Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., and Lee, K. (2006). The development of executive functioning and theory of mind: a comparison of Chinese and U.S. preschoolers. *Psychol. Sci.* 17, 74–81. doi: 10.1111/j.1467-9280.2005.01667.x
- Schuchardt, K., Maehler, C., and Hasselhorn, M. (2008). Working memory deficits in children with specific learning disorders. *J. Learn. Disabil.* 41, 514–523. doi: 10.1177/0022219408317856
- St Clair-Thompson, H. L. (2010). Backwards digit recall: a measure of short-term memory or working memory? *Eur. J. Cogn. Psychol.* 22, 286–296. doi: 10.1080/09541440902771299
- Swanson, H. L. (1993). Working memory in learning disability subgroups. *J. Exp. Child Psychol.* 56, 87–114. doi: 10.1006/jecp.1993.1027
- Swanson, H. L., and Beebe-Frankenberger, M. (2004). The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *J. Educ. Psychol.* 100, 343–379. doi: 10.1037/0022-0663.100.2.343
- Swanson, H. L., and Jerman, O. (2006). Math disabilities: a selective meta-analysis of the literature. *Rev. Educ. Res.* 76, 249–274. doi: 10.1016/j.ridd.2015.01.002
- Swanson, H. L., and Kim, K. (2007). Working memory, short-term memory, and naming speed as predictors of children's mathematical performance. *Intelligence* 35, 151–168. doi: 10.1016/j.intell.2006.07.001
- Swanson, H. L., and Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: both executive and phonological processes are important. *J. Exp. Child Psychol.* 79, 294–321. doi: 10.1006/jecp.2000.2587
- Szűcs, D., Devine, A., Soltesz, F., Nobes, A., and Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex* 49, 2674–2688. doi: 10.1016/j.cortex.2013.06.007
- Tang, W. Y. (2012). *Nonsymbolic Numerical Magnitude Processing and Arithmetic Performance: An Investigation on First-Grade Children With and Without Mathematics Difficulties*. Master's dissertation, The University of Hong Kong, Hong Kong.
- Testa, R., Bennett, P., and Ponsford, J. (2012). Factor analysis of nineteen executive function tests in a healthy adult population. *Arch. Clin. Neuropsychol.* 27, 213–224. doi: 10.1093/arclin/acr112
- Toffalini, E., Giofrè, D., and Cornoldi, C. (2017). Strengths and weaknesses in the intellectual profile of different subtypes of specific learning disorder: a study on 1,049 diagnosed children. *Clin. Psychol. Sci.* 5, 402–409. doi: 10.1177/2167702616672038
- Toll, A. D., Liu, J. M., Gulati, G., Behling, E. M., and Kocher, W. D. (2011). Does routine repeat testing of critical values offer any advantage over single testing? *Arch. Pathol. Lab. Med.* 135, 440–444. doi: 10.1043/2010-0025-OA.1
- van den Bos, K. P., Zijlstra, B. J. H., and van den Broeck, W. (2003). Specific relations between alphanumeric-naming speed and reading speeds of monosyllabic and multisyllabic words. *Appl. Psychol.* 24, 407–430. doi: 10.1017/S0142716403000213
- van der Sluis, S., de Jong, P. F., and van der Leij, A. (2004). Inhibition and shifting in children with learning deficits in arithmetic and reading. *J. Exp. Child Psychol.* 87, 239–248. doi: 10.1016/j.jecp.2003.12.002
- van der Sluis, S., de Jong, P. F., and van der Leij, A. (2005). Working memory in Dutch children with reading- and arithmetic-related LD. *J. Learn. Disabil.* 38, 207–222. doi: 10.1177/00222194050380030301
- van der Ven, S. H. G., Kroesbergen, E. H., Boom, J., and Leseman, P. P. M. (2012). The development of executive functions and early mathematics: a dynamic relationship. *Br. J. Educ. Psychol.* 82, 100–119. doi: 10.1111/j.2044-8279.2011.02035.x
- Viterbori, P., Usai, M. C., Traverso, L., and De Franchis, V. (2015). How preschool executive functioning predicts several aspects of math achievement in Grades 1 and 3: a longitudinal study. *J. Exp. Child Psychol.* 140, 38–55. doi: 10.1016/j.jecp.2015.06.014
- Vukovic, R. K., and Siegel, L. S. (2010). Academic and cognitive characteristics of persistent mathematics difficulty from first through fourth grade. *Learn. Disabil. Res. Pract.* 25, 25–38. doi: 10.1111/j.1540-5826.2009.00298.x
- Wager, T. D., Jonides, J., and Smith, E. E. (2006). Individual differences in multiple types of shifting attention. *Mem. Cognit.* 34, 1730–1743. doi: 10.3758/BF03195934
- Wang, J., and Lin, E. (2009). A meta-analysis of comparative studies on Chinese and US students' mathematics performance: implications for mathematics education reform and research. *Educ. Res. Rev.* 4, 177–195. doi: 10.1016/j.edurev.2009.06.003
- Wechsler, D. (1992). *Wechsler Individual Achievement Test*. San Antonio, TX: Psychological Corporation.
- Wechsler, H. (2009). Linguistics and face recognition. *J. Vis. Lang. Comput.* 20, 145–155. doi: 10.1016/j.jvlc.2009.01.001
- Willoughby, M. T., Magnus, B., Vernonfeagans, L., and Blair, C. B. (2016). Developmental delays in executive function from 3 to 5 years of age predict kindergarten academic readiness. *J. Learn. Disabil.* 50, 359–372. doi: 10.1177/0022219415619754
- Wise, J. C., Pae, H. K., Wolfe, C. B., Sevcik, R. A., Morris, R. D., Lovett, M., et al. (2008). Phonological awareness and rapid naming skills of children with reading disabilities and children with reading disabilities who are at risk for mathematics difficulties. *Learn. Disabil. Res. Pract.* 23, 125–136. doi: 10.1111/j.1540-5826.2008.00270.x
- Wolf, M., and Denckla, M. B. (2005). *Rapid Automatized Naming and Rapid Alternating Stimulus tests (RAN/RAS)*. Austin, TX: PRO-ED.
- Xia, Z., Zhang, L., Hoeft, F., Gu, B., Gong, G., and Shu, H. (2018). Neural correlates of oral word reading, silent reading comprehension, and cognitive subcomponents. *Int. J. Behav. Dev.* 42, 342–356. doi: 10.1177/0165025417727872
- Xue, J., Shu, H., Li, H., Li, W., and Tian, X. (2013). The stability of literacy-related cognitive contributions to Chinese character naming and reading fluency. *J. Psycholinguist. Res.* 42, 433–450. doi: 10.1007/s10936-012-9228-0

- Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., and Pieper, S. (2013). Shifting ability predicts math and reading performance in children: a meta-analytical study. *Learn. Individ. Dif.* 23, 1–9. doi: 10.1016/j.lindif.2012.10.004
- Zhang, Y., Tardif, T., Shu, H., Li, H., Liu, H., McBride-Chang, C., et al. (2013). Phonological skills and vocabulary knowledge mediate socioeconomic status effects in predicting reading outcomes for Chinese children. *Dev. Psychol.* 49, 665–671. doi: 10.1037/a0028612
- Zhao, N., Valcke, M., Desoete, A., Burny, E., and Imbo, I. (2014). Differences between Flemish and Chinese primary students' mastery of basic arithmetic operations. *Educ. Psychol.* 34, 818–837. doi: 10.1080/01443410.2013.832150

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

Copyright © 2018 Wang, Georgiou, Li and Tavouktsoglou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Association of Number and Space Under Different Tasks: Insight From a Process Perspective

Zhijun Deng¹, Yinghe Chen^{1*}, Meng Zhang², Yanjun Li¹ and Xiaoshuang Zhu¹

¹ School of Developmental Psychology, Faculty of Psychology, Beijing Normal University, Beijing, China, ² Department of Psychology, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

Samuel Shaki,
Ariel University, Israel
Krzysztof Cipora,
Universität Tübingen, Germany

*Correspondence:

Yinghe Chen
chenyinghe@bnu.edu.cn

Specialty section:

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

Received: 05 March 2018

Accepted: 24 May 2018

Published: 12 June 2018

Citation:

Deng Z, Chen Y, Zhang M, Li Y and
Zhu X (2018) The Association
of Number and Space Under Different
Tasks: Insight From a Process
Perspective. *Front. Psychol.* 9:957.
doi: 10.3389/fpsyg.2018.00957

We investigated the Spatial Numerical Association of Response Codes (SNARC) effect in 240 adults using a parity judgment and a magnitude classification task. Eight numbers from 1 to 9 except 5 were randomly presented one at a time, half of the participants were asked to compare these number with the target number 5 in the magnitude classification task; the other half of the participants were asked to judge whether these numbers were even or odd. It was called a phase when all eight numbers were tested; there were in total 16 phases. Detailed analyses of the changes in response times across the range of numbers and the chronological trend of the SNARC effect size over 16 phases estimated by a curvilinear regression model were reported. This phase-to-phase design and analyses provide an insight into the process of the SNARC effect in different tasks. We found that the SNARC effect emerged earlier and stayed more stable in magnitude classification task than in the parity task during the time course. Furthermore, the size of SNARC effect increased with time in the magnitude classification task, whereas it fluctuated up and down over time in the parity task. These findings indicate that the association of the number and space is dynamic and the process of the SNARC effect varies across tasks.

Keywords: SNARC effect, parity judgment task, magnitude classification task, phase-to-phase design and analyses, process perspective

INTRODUCTION

It is well known that the processing of numerical magnitude is closely related to spatial processing in the domain of numerical cognition (Wood et al., 2008; Fias et al., 2011). The Spatial Numerical Association of Response Codes (SNARC) effect refers to the phenomenon that individuals typically react faster to relatively smaller numbers with left-sided responses and they react faster to relatively larger numbers with right-sided responses. It is one of the most striking demonstrations of the numerical-spatial association (Dehaene et al., 1993). The SNARC effect has long been ascribed to a mental number line stored in the long-term memory (Dehaene, 1992; Gevers et al., 2003).

However, accumulating evidence suggests that many transient factors can affect The SNARC effect. For instance, the given number range and the reference number affect participants' left or right side responses (Dehaene et al., 1993; Fias et al., 1996; Ben Nathan et al., 2009): The number 5 receives faster right side responses when the overall range is 1–5, but it receives faster left side responses when the range is 4–9 (Dehaene et al., 1993). Moreover, task instructions also affect the SNARC effect (Ristic et al., 2006; Viarouge et al., 2014a): asking participants to imagine a linear

rule leads to a standard SNARC effect, whereas asking them to imagine a circular clock leads to a reversed SNARC effect (Bächtold et al., 1998). Additionally, researchers also found that the directional component of a prior spatial activity (e.g., directions in number placement or text-reading) modulated the strength of the SNARC effect (Shaki and Fischer, 2008; Fischer et al., 2010). These findings indicate that spatial-numerical associations are not fixed; they can be affected by tasks and measurements.

An impressive number of studies on spatial-numerical associations using the repetition design, which presented numerous repetitions of single digits (Wood et al., 2008; Fischer and Shaki, 2014), but only a few focused on the repetition effect. For example, in order to explore the exact task factors that affect the SNARC effect, some studies focused more precisely on the trial-to-trial changes (Notebaert et al., 2006; Fischer et al., 2010; Pfister et al., 2013). Pfister et al. (2013) examined the effect of prior trials on the SNARC effect, specifically how the preceding congruency between the target number's spatial association and the required response influenced the SNARC effect. They asked participants to perform a parity judgment task, and found that the size of SNARC effect was reduced instantly after participants experienced the preceding incongruence. Studies (Notebaert et al., 2006; Fischer et al., 2010; Pfister et al., 2013) with such sequential modulation provide a finer measurement of the dynamics of the SNARC effect, and indicate that the spatial-numerical associations could be a real-time control process. However, this trial-to-trial design may be useful for parity judgment tasks, but it can not be applied to a magnitude classification task, because in a magnitude classification task, congruent trials and incongruent trials are often separated into two blocks. The instant control over spatial-numerical associations like Pfister et al. (2013) can not be obtained in a magnitude classification task.

Both magnitude classification tasks and parity judgment tasks are commonly used methods for investigating the SNARC effect. In parity judgment tasks, participants are asked to judge whether digits are odd or even; in magnitude classification tasks, participants are asked to judge whether digits are smaller or larger than a reference number. Whether these two types of tasks involve the same processes of spatial-numerical associations are controversial.

Some studies found that the number-space associations measured by the parity task and magnitude classification task shared common processes. For example, Cheung et al. (2015) found a significant correlation between the sizes of SNARC effects in these two tasks (see also Georges et al., 2017). Furthermore, several studies suggest that the number-space associations in both parity and magnitude processing tasks arise from the verbal-spatial coding mechanisms (Gevers et al., 2010; Imbo et al., 2012). Though these findings suggest a single predominant account, accumulating evidence has indicated the task-dependent coding mechanism.

The magnitude classification task and the parity judgment task differ in many ways, therefore they may capture different aspects of spatial-numerical associations. First, the processing of magnitude information is different, magnitude information is

implicitly and automatically activated in parity judgment task, whereas it is not the case for the magnitude task, in which numerical magnitude is task-relevant and needs to be processed voluntarily (Priftis et al., 2006; Shaki and Fischer, 2018); Second, the difference also exists in the response selection stage, for magnitude classification, the same responses were associated with numbers that were smaller or larger than the referent, whereas for parity judgment, the responses alternate for each number (van Dijck et al., 2009). Besides, parity judgment has a unique effect, i.e., the MARC effect (Nuerk et al., 2004; Roettger and Domahs, 2015), where odd numbers are responded faster on left hand side and even numbers are responded faster on right hand side. This effect is usually not present in the magnitude classification task. Furthermore, studies using both tasks showed that the number-space mapping required different modalities (Herrera et al., 2008; van Dijck et al., 2009) and different amounts of working memory resources (Deng et al., 2017) for magnitude classification and parity judgment. Thus, it is desirable to explore the differences of number-space association process that is involved in magnitude classification tasks and parity judgment tasks.

The present study examined the differences in the number-space association process using the magnitude classification task and the parity judgment task. Unlike previous researchers focusing on the influence of a prior trial on the next trial (as in Notebaert et al., 2006; Fischer et al., 2010; Pfister et al., 2013), we think the process of state changes trial-to-trial, throughout all the trials (Macdonald et al., 2011; Unsworth and McMillan, 2014; Amir et al., 2016). Therefore, we adopted the trial-to-trial processing perspective and investigated the phase-to-phase changes for each participant. In our study, eight numbers from 1 to 9 (except 5) were tested in random order for both tasks; the unit of eight trials was considered as a phase. Our study was designed and data analyses (e.g., regression) were conducted in a phase-to-phase manner. We report changes of the SNARC effect during the time course of all 16 phases, where the size of the SNARC effect was represented by the regression coefficients. Our phase-to-phase design and analyses provided a micro-level perspective for better understanding the process of number and space association and its variations in different numerical tasks.

MATERIALS AND METHODS

Participants

A total of 240 native Chinese adults participated the experiment. All were right-handed and reported normal or corrected-to-normal vision. We divided participants randomly into two groups each with 120 adults: one group (74 females, 46 males; 18–28 years old, mean age 22.68 years) was assigned to complete the magnitude classification task, and the other group (73 females, 47 males; 18–29 years old, mean age 22.32 years) was assigned to complete the parity judgment task. None of the participants were familiar with the purpose of the study. We explained the procedures of the experiment and obtained participants' informed consent before experiment.

Participant each received a small amount of monetary reward after experiment.

Stimuli and Procedure

The experiment was programmed using E-Prime 2 Professional Software on a 17-in. LCD computer screen ($1,280 \times 1,024$ pixels). Stimuli were Arabic numbers (Arial font, 48 point size) in the range from 1 to 9 with the exception of 5. Between stimulus presentations, participants saw a fixation point, which was an asterisk (*) of size 48 points in the center of the screen. All stimuli were in black showing on a white background. Participants indicated their judgments by pressing either the A or L key on a standard QWERTY computer keyboard.

For both the magnitude classification task and parity judgment task, each trial started with a 300 ms presentation of the fixation asterisk, then a target number appeared in the center of the screen. Participants had to make their judgments within 5000 ms by pressing corresponding keys. In magnitude classification task, participants were asked to judge whether digits are smaller or larger than a reference number. In parity judgment task, participants were asked judge whether digits are odd or even. There would be a 1000 ms of blank screen following each trial. Participants' response accuracy and response time were recorded.

For each task, we presented a total of 128 trials (8 numbers $\times$ 16 phases) in two blocks. Each block contained 8 successive phases. In each phase, all of the 8 numbers (i.e., 1, 2, 3, 4, 6, 7, 8, and 9) were tested in a random order. These two blocks differed in their response mapping. In magnitude classification task, we had one block that mapped small numbers on the left side and large numbers on the right side, and the other block that counterbalance the mapping. In parity judgment task, we had

one block that associated even numbers with left side and odd numbers with right side, and the other block that counterbalance the association. The order of blocks was also counterbalance across participants. Before testing, participants completed six practice trials to become familiar with the procedure. Phases were labeled in the order of their occurrence, continuously numbered from the first phase of the first block to the last phase of the second block.

RESULTS

Data Treatment

We excluded trials with errors for data analyses. There were 2.29% of trials were with error for magnitude classification task; there were 3.82% of trials were with error for parity judgment task. Additionally, when participants' RT deviated from the corresponding cell mean by more than 3 standard deviations, we considered this data as outliers. There were 0.91% outliers in magnitude classification task and 1.26% outliers in parity judgment task.

Response Times

The mean RTs and standard error of the mean (SEM) of responses to each number magnitude in magnitude classification task and parity judgment task were calculated (See **Figure 1**). We performed a 2 (type of task: magnitude classification task, parity judgment task) $\times$ 8 (number magnitude: 1, 2, 3, 4, 6, 7, 8, and 9) repeated measures ANOVA on mean reaction times. The results revealed significant main effects of Task, $F(1,238) = 29.20$, $p < 0.0001$, $\eta^2 = 0.109$, and Number magnitude,

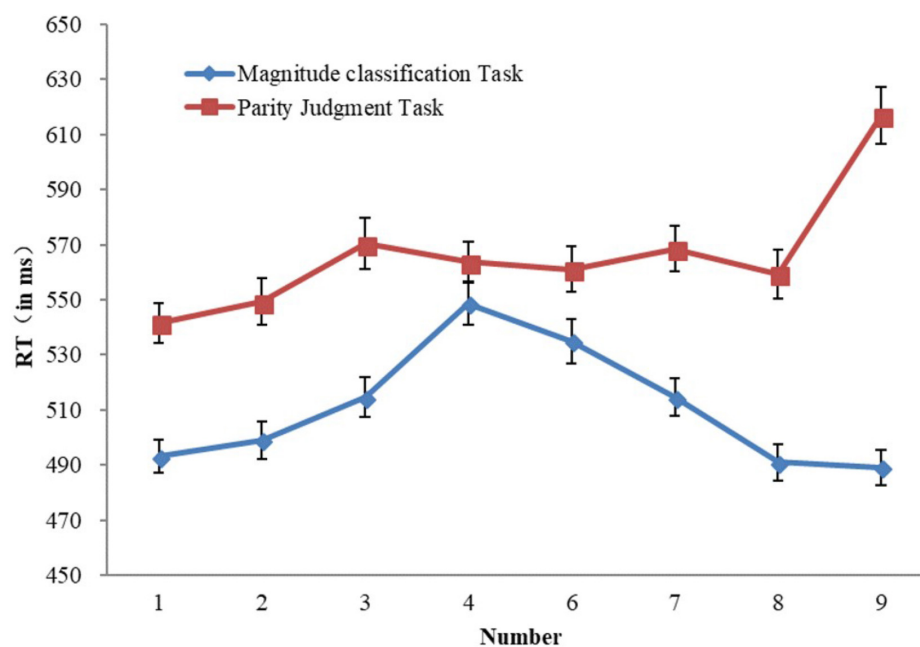


FIGURE 1 | Mean response time (RT) for magnitude classification task and parity judgment task as a function of number magnitudes.

TABLE 1 | The Means and Standard Deviations of RTs of each phase in magnitude classification task.

	Phase number															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
<i>M</i>	542.34	516.31	510.39	503.10	497.07	508.76	507.77	500.43	528.23	510.94	500.61	500.65	505.24	508.93	501.73	512.33
<i>SD</i>	145.25	127.19	125.77	123.68	110.40	132.77	120.85	123.23	148.25	132.85	125.68	114.09	121.99	124.01	122.19	133.79

TABLE 2 | The Means and Standard Deviations of RTs of each phase in parity judgment task.

	Phase number															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
<i>M</i>	605.61	575.82	560.86	565.73	559.32	558.01	559.90	563.64	573.58	552.17	568.85	564.68	564.04	567.74	546.65	560.74
<i>SD</i>	184.30	167.23	153.78	151.15	150.37	156.31	152.01	150.27	161.94	145.05	168.89	156.80	166.15	166.99	154.01	162.13

$F(7,1666) = 46.30, p < 0.0001, \eta^2 = 0.163$. The interaction was also significant, $F(7,1666) = 63.02, p < 0.0001, \eta^2 = 0.209$. Further, one way of simple effect analyses indicated that, the RTs for all the numbers in parity judgment task were significantly longer than those in magnitude classification task (all $ps < 0.05$) except for the digit 4 ($p = 0.165$). The other way of simple effects analyses suggested that, in the magnitude classification task, RTs for both 4 and 6 were significantly longer than those for any other digit (all $ps < 0.001$); RTs for 1, 2, 8, and 9 were significantly shorter than those for 3, 4, 6, and 7 (all $ps < 0.05$). In the parity judgment task, the RT for 1 was significantly shorter than those for others (all $ps < 0.001$); the RT for 9 was significantly longer than those for others (all $ps < 0.0001$).

In general, RTs for the parity judgment task were longer than those for the magnitude classification task. RTs changed across the eight number magnitudes differently in magnitude classification task and parity judgment task. In magnitude classification task, RT reflected a distance effect (Moyer and Landauer, 1967; Sekuler and Mierkiewicz, 1977), which means that RTs decreased when the distance between the standard and the target increased. In the parity judgment task, RT reflected a size effect (Starkey and Cooper, 1980; Dehaene et al., 1998), which means that RTs increased as the magnitude increases.

SNARC Effects at Individual Level and at Group Level

The SNARC effect traditionally has been indicated by the existence of a difference of response time to the same number between using left hand and using right hand, which oftentimes favors the right hand for numbers greater than 5 and the left hand for numbers less than 5.

For each participant and each number magnitude, we calculated an RT difference (dRT) for each participant by subtracting the mean RT using the left hand from the mean RT using the right hand and regressed the dRT on number magnitudes (i.e., 1–4, 6–9). The regression weights of each participant indicated their SNARC effect (Fias et al., 1996), which were used for further analyses. For both magnitude classification and parity judgment task, we examined whether the regression weights deviated significantly from zero at the group by using t -tests. For the magnitude classification task, $M = -6.25, SD = 17.84, t(119) = -3.836, p < 0.0001$. For the parity judgment task, $M = -7.59, SD = 9.92, t(119) = -8.39, p < 0.0001$. In both tasks, the slopes were significantly different from zero, indicating the presence of the SNARC effect at the group level. Moreover, an independent samples t -test was applied to compare the regression weights for magnitude classification task and parity judgment task. We found the difference was not significant, $t(238) = 0.72, p = 0.47$. The sizes of the SNARC effect in these two tasks were equal at the group level.

Additionally, as expected, the majority of participants showed the SNARC effect, which was negatively associated with the number magnitude. There were 69.2% of participants (i.e., 83) in the magnitude classification task and 75.8% of participants (i.e., 91) in the parity judgment task showed such effect.

TABLE 3 | Summary of the regression analysis for dRT in magnitude classification task with phase as a predictor.

Regression statistic	Phase number															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
B	-2.04	-4.09	-4.35	-4.96	-5.17	-4.93	-6.65	-6.05	-6.27	-7.68	-7.18	-6.63	-8.57	-10.12	-10.74	-11.38
P	0.54	0.06	0.12	0.04	0.03	0.01	0.01	0.02	0.19	0.02	0.20	0.01	0.01	0.00	0.00	0.05
R ²	0.07	0.46	0.35	0.53	0.60	0.71	0.73	0.62	0.27	0.64	0.25	0.68	0.70	0.77	0.83	0.51

TABLE 4 | Summary of the regression analysis for dRT in parity judgment task with phase as a predictor.

Regression statistic	Phase number															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
B	-13.53	-8.57	-5.58	-11.49	-8.20	-4.95	-13.26	-0.20	-1.99	-1.36	-13.78	-2.23	-7.34	-8.22	-16.67	-4.68
p	0.14	0.26	0.45	0.05	0.09	0.29	0.07	0.97	0.83	0.86	0.23	0.75	0.31	0.27	0.01	0.38
R ²	0.32	0.21	0.10	0.51	0.40	0.19	0.44	0.00	0.01	0.01	0.23	0.02	0.17	0.20	0.70	0.13

SNARC Effects Across Phases

The mean RTs and standard deviations of each phase for the parity and magnitude tasks were calculated (see **Tables 1, 2**). For the magnitude classification task, we performed repeated measures ANOVA with phase on mean reaction times. The results revealed a significant main effect of phases, $F(15,1785) = 5.594, p < 0.0001, \eta^2 = 0.045$. The *post hoc* test found that RTs for phase 1 and phase 9 were significantly longer than those for others (all $ps < 0.05$); RTs for the rest of phases did not differ from each other (all $ps > 0.05$). For the parity task, the results of repeated measures ANOVA revealed a significant main effect of phases, $F(15,1785) = 4.526, p < 0.0001, \eta^2 = 0.037$, the *post hoc* test found that the RT for phase 1 was significantly longer than those for others ($p < 0.05$); RTs for the rest of phases did not differ from each other (all $ps > 0.05$).

The SNARC effect was examined for each phase by applying regression analyses on dRT with numerical magnitudes as the predictor. For each phase, we calculated the dRT of each number by subtracting the group-level mean RT using the left hand from the group-level mean RT using the right hand. We were able to calculate dRT this way because for each phase the order of blocks was counterbalanced between participants. There were 60 participants who responded to the number with the right hand and there were other 60 participants who responded to the same number with the left hand. For all 16 phases, we calculated the group-level regression slopes as precise quantifications of the SNARC effect and R^2 as an indicator of proportion of variance explained by each regression model (Pfister et al., 2013).

As shown in **Tables 3, 4**, all 16 phases showed negative SNARC slopes for both magnitude classification task and parity judgment task. Eleven of the 16 phases in the magnitude classification task showed significantly negative regression slopes; two of 16 phases in the parity judgment task showed significantly negative regression slopes. R^2 values across all phases were comparatively low, with most less than 0.4.

For each task, we examined the chronological trend across all 16 phases by applying curve estimation, with time as an independent variable and the regression slopes of each phase as the dependent variable.

For the magnitude classification task, the size of the regression slopes increased with time, $p < 0.001$; 90.7% of variance were explained by the curvilinear regression model. For the parity judgment task, the chronological trend was not clear ($p > 0.05$) and the model only explained 0.9% of the variances (see **Table 5**).

As show in **Figure 2**, there was a growing trend of SNARC effect throughout the phases in the magnitude classification task, whereas the SNARC effect fluctuated throughout all the phases in the parity judgment task.

DISCUSSION

We investigated the SNARC effect in a parity judgment and a magnitude classification task with a relatively large sample of participants. Detailed analyses of spatial-numerical associations were reported from the perspective of processes. We observed robust SNARC effects in both the magnitude

TABLE 5 | Model summary and parameter estimates.

Equation	Model summary				Parameter estimates		
	R Square	F	df1	df2	Sig.	Constant	b1
Magnitude classification	0.907	136.724	1	14	0.000	−2.335	−0.511
Parity judgment	0.009	0.124	1	14	0.730	−8.470	0.099

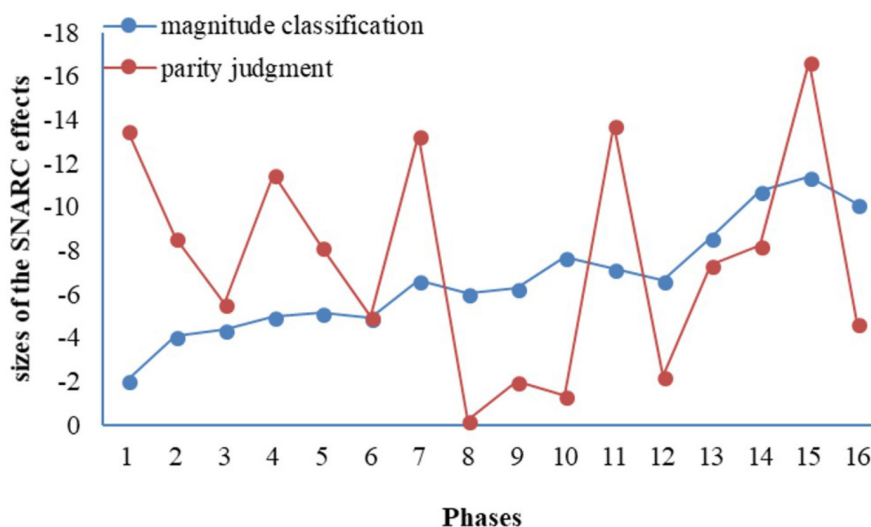
classification task and the parity judgment task, and most of participants showed negative SNARC effects (Wood et al., 2006a,b). However, analyses on RTs differences among number magnitudes and the phase-to-phase changes revealed different processes for these two tasks. These findings confirmed that the SNARC effect can be easily affected by tasks by providing evidence from the number's spatial association process.

Though both magnitude classification task and parity judgment task are widely used for exploring the SNARC effect, only a few studies focused on the repetition effect. Previous researchers (Fias et al., 1996; Viarouge et al., 2014b) found that the SNARC effect was relatively stable over sessions or blocks. In our study, we looked into more refined differences between phases within a block. With more detailed analyses regarding the time course, we provided the first evidence for chronological changes of the SNARC effect. The size of SNARC effects increased with time in the magnitude classification task, whereas in the parity task, the values of SNARC effect fluctuated up and down over time. As suggested by Pfister et al. (2013) trial-to-trial design, it is plausible when researchers zoom into the process and conduct more detailed analyses, the refined differences over the process can be observed; hence providing more information about the underlying mechanism of spatial-numerical associations. Moreover, as a repetition design our study was able to detect the temporal differences of SNARC effect also because our analyses of phases were group level and our sample size and number of

repetitions provided enough statistical power (Cipora and Nuerk, 2013; Cipora and Wood, 2017).

Furthermore, our results also showed that the SNARC effect in magnitude classification task emerged earlier and stayed more stable than it did in the parity judgment task. In the magnitude classification task, most of the sizes of SNARC effect were significantly negative and increased with time. However, in the parity task, only a few SNARC effects were significantly negative and the values fluctuated up and down over time. This task difference may be because for parity judgment task, a single phase was not long enough to establish a stable association between number and space, making the SNARC effect hard to detect. Also there was a notable dissociation between the RTs of number judgments and the values of the SNARC effect, indicating a different process in making number judgments.

The question that we may ask is why parity judgment and magnitude classification engage different processes over time. One explanation relates to differences between these two tasks, which cashed out the SNARC effect (Georges et al., 2014, 2015). In the current study, the magnitude classification task required participants to process magnitudes; they were also primed by a mental number line, especially when asked to respond to small numbers with left hand and to respond to large numbers with right hand. Whereas in the parity judgment task, the response to the task (judge whether odd or even) influences the presentation of the number-space association, therefore the number-space association for each phase was weak and unstable. The stability

**FIGURE 2 |** The SNARC effects in the magnitude classification task and parity judgment task as a function of phase.

difference between the two tasks could also explain why more working memory resources were apparently needed in the parity task than in the magnitude task (Deng et al., 2017). Working memory resources were needed to rule out the influence of the task set in the parity task, whereas they were only needed to account for the inconsistency in the magnitude task.

Overall, the process difference between the parity task and the magnitude task further illustrated that spatial-numerical interactions in implicit and explicit magnitude processing tasks potentially arise from qualitatively different cognitive mechanisms. Some studies indicated the mechanisms difference from a perspective of element analysis. For example, Georges et al. (2017) found the spatial-numerical associations (SNAs) measured by the parity task and the magnitude task correlated with individual's arithmetic performance, spatial visualization ability and visualization profile differently. van Dijck et al. (2012) found similar parity SNARC effects in normal population and patients but different magnitude SNARC effects between the two populations, indicating different origins for the two SNARC effects. Similarly, their principle analyses also extracted separate components for parity task and magnitude task, suggesting different cognitive processes were engaged. Our study showed the process where spatial-numerical associations varied in implicit and explicit magnitude processing tasks. Besides, participants' RTs in parity judgment task increased as number magnitudes increased, which corresponded to the size effect (Dehaene et al., 1998). However, their RTs in the magnitude task behaved more categorically – their pattern can be approximated by two parallel horizontal lines – one for numbers smaller than the criterion and one for numbers larger than the criterion. All these results are consistent with Gevers et al. (2006) study, thereby our study further supports the task-dependent spatial coding mechanisms (see also Wood et al., 2008).

A question that cannot be answered based on the present results is whether the differences of the SNARC effect between these two tasks reflect different number-space associations or just different task demands. Previous research pointed out that the SNARC effect was range-dependent (Dehaene et al., 1993), reference-dependent (Bächtold et al., 1998; Ristic et al., 2006), and task demand-dependent (Fischer et al., 2010; Pfister et al., 2013). These characteristics can be considered as evidence for the role of working memory in transient associations of space and number (Fias et al., 2011; De Belder et al., 2015). Alternatively, researcher (Gevers et al., 2006) adopting computational modeling argued

for a parallel activation of preexisting links between magnitude and spatial representation and short-term links created on the basis of task instructions. Recent research (Ginsburg and Gevers, 2015; Huber et al., 2016) found that the SNARC effect and the ordinal position effect resulted from the activation of different representations, which supports the computational view of number-space associations.

In conclusion, the present results trace out the process of the number and space association in a magnitude classification task and a parity judgment task. The analyses on RTs differences and the phase-to-phase changes revealed that the formation of the SNARC effect under tasks were different. These findings remind us that the type of task is also a key element in the exploration of the nature of the SNARC effect. More attention and more research need to be done to better understand the nature of SNARC effect and its variations in different tasks. To address the above questions both more empirical evidence and computational models will be helpful in the future.

ETHICS STATEMENT

This research was approved by the local ethical committee of Beijing Normal University. We obtained informed written consent from every participant, according to the institutional guidelines of Beijing Normal University.

AUTHOR CONTRIBUTIONS

ZD contributed on designing and conducting the research, on acquisition and interpretation of data, and on drafting of manuscript. YC contributed to conception and design, and on interpretation of data. MZ contributed to conception and design and drafting of manuscript. YL and XZ contributed on interpretation of data. All authors approved the final version of the manuscript for submission.

FUNDING

This study was funded by major project grants from the National Social Science Foundation of China (No. 14ZDB160) and the National Natural Science Foundation of China (No. 31271106).

REFERENCES

- Amir, I., Zvielli, A., and Bernstein, A. (2016). (de) coupling of our eyes and our mind's eye: a dynamic process perspective on attentional bias. *Emotion* 16, 978–986. doi: 10.1037/emo0000172
- Bächtold, D., Baumüller, M., and Brugger, P. (1998). Stimulus-response compatibility in representational space. *Neuropsychologia* 36, 731–735. doi: 10.1016/S0028-3932(98)00002-5
- Ben Nathan, M., Shaki, S., Salti, M., and Algom, D. (2009). Numbers and space: Associations and dissociations. *Psychon. Bull. Rev.* 16, 578–582. doi: 10.3758/PBR.16.3.578
- Cheung, C.-N., Ayzenberg, V., Diamond, R. F., Yousif, S., and Lourenco, S. F. (2015). “Probing the mental number line: a between-task analysis of spatial-numerical associations,” in *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, eds D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, et al. (Austin, TX: Cognitive Science Society.), 357–362.
- Cipora, K., and Nuerk, H. C. (2013). Is the snarc effect related to the level of mathematics? no systematic relationship observed despite more power, more repetitions, and more direct assessment of arithmetic skill. *Q. J. Exp. Psychol.* 66, 1974–1991. doi: 10.1080/17470218.2013.772215
- Cipora, K., and Wood, G. (2017). Finding the snarc instead of hunting it: a 20*20 monte carlo investigation. *Front. Psychol.* 8:1194. doi: 10.3389/fpsyg.2017.01194
- De Belder, M., Abrahamse, E., Kerckhof, E., Fias, W., and van Dijck, J. P. (2015). Serial position markers in space: visuospatial priming of serial order working memory retrieval. *PLoS One* 10:e0116469. doi: 10.1371/journal.pone.0116469

- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N
- Dehaene, S., Bossini, S., and Giraux, P. (1993). The mental representation of parity and number magnitude. *J. Exp. Psychol.* 122, 371–396. doi: 10.1037/0096-3445.122.3.371
- Dehaene, S., Dehaenelambertz, G., and Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends Neurosci.* 21, 355–361. doi: 10.1016/S0166-2236(98)01263-6
- Deng, Z., Chen, Y., Zhu, X., and Li, Y. (2017). The effect of working memory load on the SNARC effect: maybe tasks have a word to say. *Mem. Cogn.* 45, 428–441. doi: 10.3758/s13421-016-0676-x
- Fias, W., Brysbaert, M., Geypens, F., and d'Ydewalle, G. (1996). The importance of magnitude information in numerical processing: evidence from the SNARC effect. *Math. Cogn.* 2, 95–110. doi: 10.1080/135467996387552
- Fias, W., van Dijck, J. P., and Gevers, W. (2011). “How number is associated with space? The role of working memory,” in *Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought*, eds S. Dehaene and E. Brannon (New York, NY: Elsevier), 133–148.
- Fischer, M. H., Mills, R. A., and Shaki, S. (2010). How to cook a SNARC: Number placement in text rapidly changes spatial–numerical associations. *Brain Cogn.* 72, 333–336. doi: 10.1016/j.bandc.2009.10.010
- Fischer, M. H., and Shaki, S. (2014). Spatial associations in numerical cognition—from single digits to arithmetic. *Q. J. Exp. Psychol.* 67, 1461–1483. doi: 10.1080/17470218.2014.927515
- Georges, C., Hoffmann, D., and Schiltz, C. (2014). Cognitive style influences number-space associations. *Int. Conf. Trans. Optic. Netw.* 29, 1–4.
- Georges, C., Hoffmann, D., and Schiltz, C. (2017). How and why do number-space associations co-vary in implicit and explicit magnitude processing tasks. *J. Num. Cogn.* 3, 182–211. doi: 10.5964/jnc.v3i2.46
- Georges, C., Schiltz, C., and Hoffmann, D. (2015). Task instructions determine the visuospatial and verbal-spatial nature of number-space associations. *Q. J. Exp. Psychol.* 68, 1895–1909. doi: 10.1080/17470218.2014.997764
- Gevers, W., Lammertyn, J., Notebaert, W., Verguts, T., and Fias, W. (2006). Automatic response activation of implicit spatial information: evidence from the snarc effect. *Acta Psychol.* 122, 221–233. doi: 10.1016/j.actpsy.2005.11.004
- Gevers, W., Reynvoet, B., and Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition* 87, B87–B95. doi: 10.1016/S0010-0277(02)00234-2
- Gevers, W., Santens, S., Dhooge, E., Chen, Q., Van den Bossche, L., and Fias, W. (2010). Verbal-spatial and visuospatial coding of number–space interactions. *J. Exp. Psychol.* 139, 180–190. doi: 10.1037/a0017688
- Ginsburg, V., and Gevers, W. (2015). Spatial coding of ordinal information in short-and long-term memory. *Front. Hum. Neurosci.* 9:8. doi: 10.3389/fnhum.2015.00008
- Herrera, A., Macizo, P., and Semenza, C. (2008). The role of working memory in the association between number magnitude and space. *Acta Psychol.* 128, 225–237. doi: 10.1016/j.actpsy.2008.01.002
- Huber, S., Klein, E., Moeller, K., and Willmes, K. (2016). Spatial–Numerical and ordinal positional associations coexist in parallel. *Front. Psychol.* 7:438. doi: 10.3389/fpsyg.2016.00438
- Imbo, I., Brauwer, J. D., Fias, W., and Gevers, W. (2012). The development of the snarc effect: evidence for early verbal coding. *J. Exp. Child Psychol.* 111, 671–680. doi: 10.1016/j.jecp.2011.09.002
- Macdonald, J. S., Mathan, S., and Yeung, N. (2011). Trial-by-trial variations in subjective attentional state are reflected in ongoing prestimulus EEG alpha oscillations. *Front. Psychol.* 2:82. doi: 10.3389/fpsyg.2011.00082
- Moyer, R. S., and Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature* 215, 1519–1520. doi: 10.1038/2151519a0
- Notebaert, W., Gevers, W., Verguts, T., and Fias, W. (2006). Shared spatial representations for numbers and space: the reversal of the SNARC and the Simon effects. *J. Exp. Psychol.* 32, 1197–1207. doi: 10.1037/0096-1523.32.5.1197
- Nuerk, H. C., Iversen, W., and Willmes, K. (2004). Notation modulation of the snarc and the marc (linguistic markedness of response codes) effects. *Q. J. Exp. Psychol. A* 57, 835–863. doi: 10.1080/02724980343000512
- Pfister, R., Schroeder, P. A., and Kunde, W. (2013). SNARC struggles: Instant control over spatial–numerical associations. *J. Exp. Psychol.* 39, 1953–1958. doi: 10.1037/a0032991
- Priftis, K., Zorzi, M., Meneghello, F., Marenzi, R., and Umiltà, C. (2006). Explicit versus implicit processing of representational space in neglect: dissociations in accessing the mental number line. *J. Cogn. Neurosci.* 18, 680–688. doi: 10.1162/jocn.2006.18.4.680
- Ristic, J., Wright, A., and Kingstone, A. (2006). The number line effect reflects top-down control. *Psychon. Bull. Rev.* 13, 862–868. doi: 10.3758/BF03194010
- Roettger, T. B., and Domahs, F. (2015). Grammatical number elicits snarc and marc effects as a function of task demands. *Q. J. Exp. Psychol.* 68, 1231–1248. doi: 10.1080/17470218.2014.979843
- Sekuler, R., and Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Dev.* 48, 630–633. doi: 10.2307/1128664
- Shaki, S., and Fischer, M. H. (2008). Reading space into numbers—a cross-linguistic comparison of the SNARC effect. *Cognition* 108, 590–599. doi: 10.1016/j.cognition.2008.04.001
- Shaki, S., and Fischer, M. H. (2018). Deconstructing spatial-numerical associations. *Cognition* 175, 109–113. doi: 10.1016/j.cognition.2018.02.022
- Starkey, P., and Cooper, R. G. (1980). Perception of numbers by human infants. *Science* 210, 1033–1035. doi: 10.1126/science.7434014
- Unsworth, N., and McMillan, B. D. (2014). Trial-to-trial fluctuations in attentional state and their relation to intelligence. *J. Exp. Psychol.* 40, 882–891. doi: 10.1037/a0035544
- van Dijck, J. P., Gevers, W., and Fias, W. (2009). Numbers are associated with different types of spatial information depending on the task. *Cognition* 113, 248–253. doi: 10.1016/j.cognition.2009.08.005
- van Dijck, J. P., Gevers, W., Lafosse, C., and Fias, W. (2012). The heterogeneous nature of number-space interactions. *Front. Hum. Neurosci.* 5:182. doi: 10.3389/fnhum.2011.00182
- Viarouge, A., Hubbard, E. M., and Dehaene, S. (2014a). The organization of spatial reference frames involved in the SNARC effect. *Q. J. Exp. Psychol.* 67, 1484–1499. doi: 10.1080/17470218.2014.897358
- Viarouge, A., Hubbard, E. M., and McCandliss, B. D. (2014b). The cognitive mechanisms of the snarc effect: an individual differences approach. *PLoS One* 9:e95756. doi: 10.1371/journal.pone.0095756
- Wood, G., Nuerk, H. C., and Willmes, K. (2006a). Crossed hands and the snarc effect: a failure to replicate dehaene, Bossini and Giraux (1993). *Cortex* 42, 1069–1079.
- Wood, G., Nuerk, H. C., and Willmes, K. (2006b). Variability of the SNARC effect: systematic interindividual differences or just random error? *Cortex* 42, 1119–1123. doi: 10.1016/S0010-9452(08)70223-5
- Wood, G., Willmes, K., Nuerk, H. C., and Fischer, M. H. (2008). On the cognitive link between space and number: a meta-analysis of the SNARC effect. *Psychol. Sci. Q.* 50, 489–525.

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

Copyright © 2018 Deng, Chen, Zhang, Li and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Counting and Number Line Trainings in Kindergarten: Effects on Arithmetic Performance and Number Sense

Ilona Friso-van den Bos^{1*}, Evelyn H. Kroesbergen² and Johannes E. H. Van Luit¹

¹ Department of Special Education, Cognitive & Motor Disabilities, Utrecht University, Utrecht, Netherlands, ² Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

OPEN ACCESS

Edited by:

Annemie Desoete,
Ghent University, Belgium

Reviewed by:

Koen Luwel,
KU Leuven, Belgium
Cristina Semeraro,
Università degli Studi di Bari Aldo
Moro, Italy

*Correspondence:

Ilona Friso-van den Bos
i.vandenbos@uu.nl

Specialty section:

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

Received: 23 February 2018

Accepted: 25 May 2018

Published: 19 June 2018

Citation:

Friso-van den Bos I, Kroesbergen EH
and Van Luit JEH (2018) Counting
and Number Line Trainings
in Kindergarten: Effects on Arithmetic
Performance and Number Sense.
Front. Psychol. 9:975.
doi: 10.3389/fpsyg.2018.00975

Children's early numerical capacities form the building blocks for later arithmetic proficiency. Linear number placements and counting skills are indicative of mapping, as an important precursor to arithmetic skills, and have been suggested to be of vital importance to arithmetic development. The current study investigated whether fostering mapping skills is more efficient through a counting or a number line training program. Effects of both programs were compared through a quasi-experimental design, and moderation effects of age and socio-economic status (SES) were investigated. Ninety kindergartners were divided into three conditions: a counting, a number line, and a control condition. Pretests and posttests included an arithmetic (addition) task and a battery of number sense tasks (comparison, number lines, and counting). Results showed significantly greater gains in arithmetic, counting, and symbolic number lines in the counting training group than in the control group. The number line training group did not make significantly greater gains than the control group. Training gains were moderated by age, but not SES. We concluded that counting training improved numerical capacities effectively, whereas no such improvements could be found for the number line training. This suggests that only a counting approach is effective for fostering number sense and early arithmetic skills in kindergarten. Future research should elaborate on the parameters of training programs and the consequences of variation in these parameters.

Keywords: number sense, counting, number line, training, children, arithmetic

INTRODUCTION

Children's early numerical capacities have received growing interest in the past decade: numerical skills in kindergarten form the building blocks for later proficiency in mathematics (e.g., Passolunghi and Lanfranchi, 2012; Hornung et al., 2014). Number sense is the term most often used to characterize the intuitive understanding of number and quantities and their relations (Dehaene, 1997; Gersten and Chard, 1999; Spelke, 2000). Number sense refers to a cognitive framework that allows a child to understand, for example, the difference between having two or three sweets, but gradually develops into a much more advanced system of conceptual knowledge that allows a person to intuitively understand abstract number relations and algorithms. Various skills have been thought to be at the root of number sense development, among which the ability to

map between symbolic and non-symbolic magnitudes (Dehaene, 2001; Mundy and Gilmore, 2009; Desoete et al., 2012; Kolkman et al., 2013). In the current study, this mapping ability is trained in typically performing kindergartners using two different training programs, in order to investigate how the skill is best fostered and how arithmetic skills can be fostered through mapping.

Kindergarten number sense can be divided into three skills: non-symbolic skills, symbolic skills, and mapping between non-symbolic and symbolic skills. Mapping has been found to be the strongest predictor of later mathematical performance, and was hypothesized to restructure non-symbolic number knowledge of a child into a more conventional cognitive concept of mathematics (Kolkman et al., 2013). Mapping is considered to lie at the root of adequate development of arithmetic skills (Siegler and Booth, 2004; Booth and Siegler, 2008; Wong et al., 2016) and refers to a flexible integration between non-symbolic and symbolic quantity processing, meaning that children with well-developed mapping skills are able easily to transcode between number words, number symbols, and non-symbolic quantities. This transcoding ability may also make symbolic quantities more meaningful to children, which is essential for adequate arithmetic development (Wong et al., 2016).

Two lines of enquiry focus on the formation of mapping skills in young children, the first of which focuses on counting skills. Counting is considered a prerequisite for forming links between symbolic and non-symbolic processing (Lipton and Spelke, 2005; Le Corre and Carey, 2007). Reciting the counting sequence may help children understand the cardinal value of number words, thereby realizing that each number word relates to an exact quantity using bottom-up processing (Noël and Rousselle, 2011). In bottom-up processing, the individual stimulus, in this case the quantity or number, is used to construct an understanding of a system as a whole, in this case a system of numbers and their quantitative relations such as bigger and smaller numbers. A second line of enquiry focuses on linear placements of numbers on a number line, which are indicative of mapping skills. Acuity on a number line task is predictive of mathematics performance (Booth and Siegler, 2008; Schneider et al., 2018), and can be fostered in a top-down processing framework through number line activities, such as playing numerical linear board games (Siegler and Ramani, 2009; Fisher et al., 2011; Dackermann et al., 2016), thereby forming a novel approach to intervention in numerical skills. In this top-down presentation of number relations, numbers are understood through their placement on a scale with a predetermined number range, which forms the context for the task, and the scale itself must be understood before individual items can be placed on the scale (the number line).

Counting

Knowledge of counting and number symbols is considered an important predictor of arithmetic performance (Kolkman et al., 2013). Counting skills could predict the extent to which children can estimate numerosities (Lipton and Spelke, 2005) and place numbers on an empty number line (Desoete et al., 2013; Simms et al., 2013; Friso-van den Bos et al., 2014). It has been proposed that finger counting helps children associate between symbolic magnitudes and non-symbolic sets of items through the finger

pattern, as well as understand basic operations such as addition (Noël, 2005; Gracia-Bafalluy and Noël, 2008; Moeller et al., 2011) using a bottom-up process in which combining small numbers of objects into a bigger unit developmentally precedes more complex operations with bigger numbers. Nearly all number sense trainings include practicing counting procedures and knowledge of number symbols (e.g., Van Luit and Schopman, 2000; Krajewski et al., 2008; Kroesbergen et al., 2012; Toll and Van Luit, 2014), and isolated practice of counting procedures has been found to generalize to improved multiplication proficiency (Blöte et al., 2006). It was suggested that mapping, as the most important factor of number sense, develops through counting skills, as described by Le Corre and Carey (2007), who postulated that children make analogies between the sequence in the count list and quantifiable sets of objects, and use induction to learn to understand the correspondence between the addition of an item to a set and the progression through the count list. This implies that the mapping between number words and tangible quantities is first understood by a child through the bottom-up process of counting, making counting a first step toward a more abstract concept of number.

Number Lines

Number line placement acuity is also thought important for the development of both arithmetic skill and broader mathematical skills (Geary et al., 2007; Booth and Siegler, 2008; Schneider et al., 2018). In a number line task, a child places a target number on an empty number line bounded by the begin- and endpoints marked on either side of the line – a top-down approach in which the number range has been framed and individual units need to be placed within this framework. To use number lines in number tasks, children need to be able to relate a number to the corresponding quantity and consequently realize that a number obtains its position on the number line through its quantitative value. Typically, young children make non-linear placements, adhering to a logarithmic or power model of placements, while older children show a more linear pattern of number placement, with equal spacing between numbers of various sizes (Siegler and Opfer, 2003; Siegler and Booth, 2004; Booth and Siegler, 2006; Barth and Paladino, 2011). A more linear pattern of placements is predictive of higher achievement in arithmetic in children (Booth and Siegler, 2008). Acuity of number line placements may be interpreted as a child's ability to map between symbolic numbers and non-symbolic quantities (Kolkman et al., 2013). The symbolic numbers, in this interpretation, are the numbers to be placed on the empty number line, and the non-symbolic quantity is represented as the continuous space between the extremities of the number line.

The training of number line placement has received growing interest in the past few years (e.g., Fisher et al., 2011; Ramani and Siegler, 2011; Dackermann et al., 2016). Playing numerical linear board games, in which linear ordering of numbers was emphasized, has repeatedly been reported to improve successfully number line acuity (Siegler and Ramani, 2009; Fisher et al., 2011) and thereby facilitate the mapping between numbers and quantities. Furthermore, placement of numbers along the continuum of a number line may be seen as a form of visual

imagery of number information; a prerequisite for successful acquisition of the more complex algorithm skills needed in advanced mathematics (Zhang et al., 2012). Number line training has also been demonstrated to enhance arithmetic performance in a study among kindergarten children, but there was no transfer effect to other measures of number sense (Maertens et al., 2016).

The Current Study

The present study aimed to investigate whether the development of number sense and consequent arithmetic skills is more efficient through a bottom-up counting or a top-down number line training program in comparison to a control group through experimental training studies. We expected both trainings to have significant effects on measures of arithmetic and mapping, and small transfer effects on symbolic number sense measures. Moreover, we expected children enrolled in a number line training to make greater gains on a measure of number line acuity and children enrolled in a counting training to make greater gains on a counting task than other groups, because these skills were directly trained in these groups. However, we did not expect significant training gains on non-symbolic number sense measures because previous research showed no direct relations between non-symbolic number sense and mapping, and relations between arithmetic and number sense were dominated by mapping skills rather than non-symbolic number sense (Kolkman et al., 2013). Measures of non-symbolic number sense were included nevertheless to get a full account of training effects on number sense. Gains made after the interventions may reflect the way in which kindergarten children normally (without intervention) construct number knowledge, because according to Piaget's theory of cognitive development (Piaget, 1970), an intervention that fosters knowledge the way children intuitively approach it is likely to have greater effect than an intervention that teaches children to think differently about the matter at hand than they intuitively do.

Intervention in number sense skills in children of low socio-economic status (SES) has aroused specific research interest (Siegler and Ramani, 2008; Jordan et al., 2012; Dyson et al., 2013). Greater gains have been reported for children from a low SES background than for children from middle SES backgrounds (Starkey et al., 2004). The current study attempted to replicate this finding by creating a distinction between children from high or low to average SES and investigating whether training is equally effective for both groups of children. Finally, because age has been found to explain differences in intervention outcomes between studies (Kroesbergen and Van Luit, 2003), the age of the children was included as a moderator variable.

MATERIALS AND METHODS

Participants

Ninety Dutch kindergartners with a mean age of 5 years and 8 months ($SD = 4$ months; range: 5;0–6;6 years) participated in the study. Data of one child were removed because his data produced outliers on multiple variables. Of the remaining children, 47 were girls (52.8%). The number of children born in

the Netherlands was 83 (93.2%). Of all children born outside of the Netherlands, at least one parent was born in the Netherlands.

In each class, nine children participated. After pretesting, children were divided into three conditions: (1) a counting training condition, (2) a number line training condition, and (3) an 'education as usual' control group. They were distributed across conditions in such a way that their counting and arithmetic scores at pretest were approximately equal between groups, with three children from each class participating in each condition, to regulate group size and prevent differences in key outcome measures at pretest. There was no age difference between the three groups, $F(2,86) = 0.55$, $p = 0.58$, nor was there a difference between any of the groups in any of the outcome measures during pretest (ps ranging from 0.44–0.97).

Socio-economic status was measured using short questionnaires filled out by parents. Children were coded as coming from families with high SES if they had at least one parent who had completed higher education. Of the children, 56 were coded as being from high SES families, 32 as being from low to average SES families, and for one child no data were available. Children from both SES categories were distributed equally across training conditions, $\chi^2(2, N = 88) = 0.22$, $p = 0.89$.

Interventions

Each intervention group received 12 training sessions spread over 6 weeks in groups of three children, lasting approximately 20 min per session. Difficulty of the sessions increased by extending the range of numbers included in the games: numbers up to 10 were included in the first four sessions, numbers up to 20 in the next four sessions, and numbers up to 50 in the last four sessions. This range was especially included because most children at the end of kindergarten already know the range up to 20. In both training programs, four games were played in total, and two per session, so that each game was played every other session. Number cards were used to support the activities in each training program. Training sessions were not specifically planned during class mathematics activities.

Counting Training

The counting training consisted of the following activities, presented as games:

1. Resultative counting, using various simple motor activities such as clapping, colorful cards, and objects to count. Counting skills are predictive of proficiency in other number-related tasks (Lipton and Spelke, 2005).
2. Counting on from a number higher than 1. Various sets of colorful cards were used for this activity. Counting on is a counting activity that requires the understanding of cardinality (Van Luit and Van de Rijt, 2009).
3. Non-linear board game: in this board game, a six-sided die was used to indicate the moves the children could make. The squares on the board were not numbered, and some of them contained an icon initiating a counting challenge. Playing non-linear board games is not expected to facilitate number line acuity, as opposed to linear board games (Ramani and Siegler, 2011).

- Counting on: colorful stones of the same shape and size were added to a pillow case to stimulate the use of shortened counting and illustrate the concept of addition.

Number Line Training

The number line training consisted of the following activities, presented as games:

- Number-to-position game: an empty number line in the form of a printed test tube was presented on worksheets, and a number was located on the number line by the children with a pencil. This game was based on the number-to-position task (Siegler and Booth, 2004; Laski and Siegler, 2007). For each session, ten numbers were chosen that covered the entire range of the number line, and the order of the items was semi-randomized.
- Position-to-number game: a number line with a given position was presented on worksheets, and the children were asked to assign a number to the location. This activity was based on the number line task (Siegler and Booth, 2004; Laski and Siegler, 2007), but the position was the information given to the child, and the number the response by the child. The beginning and endpoint were given. Target numbers covered the entire range of the number line and were presented in semi-random order.
- Linear numerical board game (Siegler and Ramani, 2008): in this board game, a six-sided die was used to indicate the moves the children could make. The squares on the board were numbered, and the child was asked to state out loud which squares were crossed during a move.
- Linear numerical tag game: in this game, children simulated a tag game on a board with numbered squares. They chose a starting position and took turns to roll a die indicating the number of steps (one to six) they could move in either direction of the number line. They were challenged to try to locate their tokens on the same square as another child and received points for each time this happened. The children were asked to state out loud which squares were crossed during their move.

Control Group

A control group received education as usual and did not participate in any research-related extra activities. Children in the Netherlands typically receive full-day programs from the day they turn 4 years old. Mathematics is part of every kindergarten curriculum, and is taught through various age-appropriate activities.

Measures

Arithmetic

To measure early arithmetic proficiency, children completed 16 addition problems displayed on the laptop screen. Of the problems, 11 had an answer below 10 (e.g., $5 + 3$) and 5 had an answer between 10 and 20 (e.g., $6 + 8$). Tie problems were not included in the set of items. All problems were preceded by a 2-s alerting phase. The score was the number of correct answers. Internal consistency at pretest was high, $\alpha = 0.94$.

Number Sense: Comparison Tasks

The comparison task had two versions. In the symbolic version, participants judged which of two Arabic numbers was bigger than the other through a key press using the hand corresponding to the location (left or right) of the selected stimulus on the screen. All numbers were between 1 and 9. Each trial was preceded by an alerting beep, and 1500 ms after the beep, the stimuli appeared. The maximum response time was 5 s. There were four practice trials and 26 test trials, and total accuracy was used as the score of the child. Numerical distance could range from 1 to 4, each distance appearing 8, 7, 6, and 5 times, respectively. The largest number appeared on both sides of the screen 13 times. Symbolic comparison tasks can be seen as measures of mapping because it has been hypothesized that children use the mental number line to complete the task (see: Kolkman et al., 2013). Test-retest reliability of a similar task has been found to be good (Clarke and Shinn, 2004).

The non-symbolic version of the comparison task was mostly the same as the symbolic version, but sets of dots appeared instead of Arabic numbers, controlled for dot size and surface array, and counterbalanced for the location of the correct response. To prevent counting strategies, the stimuli disappeared from the screen after 840 ms. Numerical distance could range from 1 to 4, each distance appearing 8, 7, 6, and 5 times, respectively, with a total of 26 trials. All trials were preceded by an alerting beep, and 1500 ms after the beep, the stimuli appeared. The maximum response time was 5 s. There were four practice trials and 26 test trials, and total accuracy was used as the score of the child. Numerical distance could range from 1 to 4. Non-symbolic comparison tasks can be seen as measures of non-symbolic NS.

Number Sense: Number Line Tasks

The number line task (Siegler and Opfer, 2003) had two versions. In the symbolic version, Arabic numbers between 1 and 100 were displayed onscreen below a horizontal line. On both sides of the line, the minimum (1) and maximum (100) were given, and participants pointed to the position on the line they selected for the target number. Twenty-two test trials were presented to the participants, preceded by two practice trials in which they located the numbers 1 and 100 on the line and received feedback. Test trials were the numbers 2, 4, 9, 11, 14, 17, 23, 26, 31, 38, 44, 45, 52, 59, 61, 66, 73, 78, 84, 86, 92, and 99. Trials were presented in random order. No feedback was given during the testing phase. Symbolic number line tasks can be seen as measures of mapping (Kolkman et al., 2013). The score of the child was the explained variance of a linear slope (R^2), indexed by the squared correlation between estimated and actual positions.

The non-symbolic version of the number line task was similar to the symbolic version, the only difference being that the children located arrays of dots on the number line. We did not control for any visual properties of these dots such as size or surface array. Minimum and maximum were displayed in non-symbolic form as well. The same numbers were used as in the symbolic version, and numbers were also presented in random order. The score of the child was the explained variance of a linear slope (R^2). Non-symbolic number line tasks can be seen as measures of non-symbolic NS.

Number Sense: Counting

To measure the counting skills, subscales of the ENT-R (Van Luit and Van de Rijt, 2009) were used. The original ENT-R consists of nine subscales. In this study, only the subscales that measure counting were used, namely: (1) Use of number words, such as rote counting; (2) Structured counting, such as counting in two's; (3) Resultative counting, such as counting out a set of objects; and (4) General understanding of number words, such as indicating which whole number is exactly between two other numbers. Each subscale contains five items with counting tasks ranging up to 20. Resultative counting up to 20 is expected of children at the end of kindergarten. One point was awarded for each correct answer. Internal consistency of this test is good (Van Luit and Van de Rijt, 2009). This scale can be seen as a measure of symbolic number sense (Kolkman et al., 2013).

Procedure

The current study was part of the MathChild study, which was funded with a project grant from the Netherlands Organisation for Scientific Research (NWO), grant 411-07-113. The study proposal was evaluated for both quality and ethical standards, and approved by NWO. The study conformed to national and international standards of ethical research, as summarized in the Netherlands Code of Conduct for Academic Practice (Association for Universities in the Netherlands, 2004). Active parental consent was obtained prior to data collection.

Pretests and posttests were conducted individually with an interval of 6–8 weeks. The children were tested in a quiet room inside the school by undergraduate students. All tasks were administered on a laptop computer using E-Prime 1.2 software (Psychological Software Tools¹). Because of budget limitations, a variety of laptop computer brands and models was used, therefore screen sizes varied as well. Prior to testing, the students administering the tests were trained in the use of the software and the standardized instruction and registration of the tasks in a 2-h group session and successive self-guided practice exercises covering all the instruments. The pretest and posttest tasks were divided into two sessions, which took place on 2 days no more than 1 week apart. After each session, children were rewarded with a colorful sticker. During the first session, children completed working memory tasks (not included in the current analyses), arithmetic, symbolic and non-symbolic comparison, and during the second session, children completed the ENT, the symbolic number line task, and the non-symbolic number line task. No variations in task order were made.

Training sessions took place inside the school, in groups of three children, and were led by trained undergraduate students. Sessions were planned with the teacher and conducted by the undergraduate students. The training sessions were not digitalised, but conducted using colorful materials such as play boards and pawns. Posttesting took place no more than 2 weeks after the last session.

¹<http://www.pstnet.com>

Analytical Strategy

To address the research questions, Hierarchical Linear Modeling (HLM) was applied using the software package HLM version 6.06. Scores on the various tasks measuring arithmetic proficiency and number sense served as dependent variables. Three-level hierarchical models were estimated with measurement occasion at the first level, individual children at the second level, and the groups in which children were trained at the third level (children in the control group were nested in kindergarten classes). Main effects of occasion (level 1) and training condition (level 3) were added first, and interactions between occasion and training condition were added in a second step, indicating differential growth of children in each of the conditions. If significant interactions were found, *post hoc* analyses using the counting training as a reference category were conducted to investigate the difference between children in the counting and number line condition.

To control for Type I errors, the Benjamini–Hochberg correction was used, in which alpha values are adjusted for the number of analyses reported (in this case: six hierarchical regression analyses, or one analysis for each of the measures listed) based on the rank-order of *p*-values (Benjamini and Hochberg, 1995). Probability values are not compared with a static alpha value, but with a corrected ' α '. Separate corrections were performed for *post hoc* analyses.

RESULTS

Correlations between measures at pretest and moderators can be found in **Table 1**. Descriptive statistics of all three groups and the total sample in each measure can be found in **Table 2**.

Training Effects

Results of the hierarchical regression analyses are reported in **Table 3** for all measures. Each model concerns one of the outcome measures. The variable Time is indicative of mean growth across all conditions between pretest and posttest. Analyses show that mean growth between pretest and posttest was significant for counting and for non-symbolic number line performance, but not for any other measure (**Table 3**).

The interactions between time and condition are indicative of divergence in growth between the experimental group and the control group, the latter group serving as a reference group. Results show that arithmetic scores were significantly predicted by an interaction between counting training and time, indicative of larger gains within the counting group (explained variance at the occasion level: 17.85%; see **Table 3**). There was no evidence for greater gains within the number line group than in the control group. *Post hoc* tests indicated that the counting group did not make greater progress than the number line group, $B = -1.26$, $\beta = -0.09$, $p = 0.11$.

There was no interaction between training condition and time on the symbolic comparison test (explained variance at the occasion-level: 3.66%; see **Table 3**). Interaction effects on scores of non-symbolic comparison were not significant either (explained variance at the occasion level: 2.56%). No *post hoc*

TABLE 1 | Correlations between measures at pretest, and moderator variables.

	1	2	3	4	5	6	7
(1) Symbolic comparison	–						
(2) Non-symbolic comparison	0.31**	–					
(3) Symbolic number line	0.33**	0.34**	–				
(4) Non-symbolic number line	0.23*	0.36**	0.45***	–			
(5) Counting	0.38***	0.37***	0.58***	0.47***	–		
(6) Arithmetic	0.42***	0.37***	0.58***	0.51***	0.72***	–	
(7) Age	0.03	0.13	0.12	0.18	0.23*	0.25*	–
(8) SES	0.11	0.22*	0.05	0.02	0.22*	0.19	

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; all correlations are Pearson's correlations, except correlations with SES, which are Spearman's rank correlations.

TABLE 2 | Descriptive statistics at pretest and posttest, for the counting group, number line group, control group, and total sample.

Task	Counting group		Number line group		Control group		Total	
	Pretest mean (SD)	Posttest mean (SD)	Pretest mean (SD)	Posttest mean (SD)	Pretest mean (SD)	Posttest mean (SD)	Pretest mean (SD)	Posttest mean (SD)
Arithmetic	6.03 (5.67)	7.93 (4.47)	5.93 (5.46)	6.77 (5.42)	6.85 (5.54)	7.14 (4.97)	6.25 (5.51)	7.27 (4.94)
Symbolic comparison	20.67 (4.77)	20.97 (4.49)	20.77 (5.06)	21.83 (3.99)	21.28 (4.41)	22.10 (2.80)	20.90 (4.71)	21.63 (3.82)
Non-symbolic comparison	20.83 (4.46)	21.43 (3.76)	20.93 (4.70)	20.48 (3.92)	21.10 (3.50)	21.52 (3.44)	20.96 (4.21)	21.14 (3.70)
Symbolic number line*	0.32 (0.25)	0.47 (0.23)	0.35 (0.26)	0.45 (0.29)	0.40 (0.27)	0.41 (0.30)	0.36 (0.26)	0.44 (0.27)
Non-symbolic number line	0.55 (0.22)	0.66 (0.20)	0.54 (0.26)	0.65 (0.22)	0.48 (0.28)	0.63 (0.27)	0.52 (0.25)	0.65 (0.23)
Counting (ENT-R)	11.13 (4.98)	14.43 (4.01)	11.50 (4.69)	13.83 (4.29)	11.79 (4.04)	13.55 (4.14)	11.47 (4.55)	13.94 (4.12)

*Scores on the number line tasks reflect fit with a linear trend of individual data points. When using median estimates of all children on the symbolic number line task, fit with a linear and a logarithmic trend at pretest was comparable to previously reported estimates of fit (Berteletti et al., 2010): $R^2_{lin} = 0.74$, $R^2_{log} = 0.96$. Moreover, data of all except two children correlated positively with the presented numbers, indicating that the children understood the task well.

analyses were conducted for either symbolic or non-symbolic comparison.

There was a significant interaction between training condition and time predicting scores on the symbolic number line test, indicative of larger gains in the group following the counting training in comparison to the control group, but not the number line training (explained variance at the occasion level: 21.50%; see Table 3). *Post hoc* analyses indicated that the counting group did not make more gains than the number line group, $B = -0.07$, $\beta = -0.09$, $p = 0.25$. There was no significant effect of an interaction between training condition and time on non-symbolic number line performance (explained variance at the occasion level: 34.68%; see Table 3).

Finally, in the model predicting counting (ENT-R) scores, there was a significant interaction between counting training and time (Table 3), but not between number line training and time, indicative of greater gains within the counting group, but not the number line group, than in the control group (explained variance at the occasion level: 45.43%). *Post hoc* tests indicated that the counting group did not progress more than the number line group $B = -0.81$, $\beta = -0.06$, $p = 0.45$.

Moderation of SES and Age

For all measures in which there was divergence in growth between the experimental groups and the control group, main effects of SES and age were included, as well as interactions between these variables and training gains, to investigate whether

change in scores in number sense and mathematics could be explained by variation in SES and/or age of the children. Significant interaction effects were indicative of divergence in growth between children of various SES or ages.

The SES did not predict growth in any of the measures for the children enrolled in the counting training, or for arithmetic, symbolic number line, or counting scores for children enrolled in either of the training groups (all $ps > 0.05$). Age of the children predicted growth of children enrolled in the counting training in arithmetic, $B = -0.74$, $\beta = -4.57$, $p = 0.03$, and symbolic number line scores, $B = -0.04$, $\beta = -4.78$, $p < 0.01$, but not in counting, $B = -0.45$, $\beta = -3.21$, $p = 0.07$, and it did not predict growth in scores of children enrolled in the number line training, all $ps > 0.05$.

CONCLUSION AND DISCUSSION

In the current study, the possibilities of advancing number sense and arithmetic using bottom-up counting training and top-down number line training were investigated. Both counting skills and number line skills may be used to fine-tune mapping between symbolic and non-symbolic representations (Le Corre and Carey, 2007; Booth and Siegler, 2008; Noël and Rousselle, 2011), which may form a foundation for arithmetic development (Wong et al., 2016). The current study investigated mapping in a quasi-experimental design. We attempted to foster mapping skills using counting activities and number line activities, both of

TABLE 3 | Hierarchical linear regression models for task performance.

	Arithmetic					Symb. comparison					Non-symb. comparison				
	B	β	df	p	α	B	β	df	p	α	B	β	df	p	α
Model intercept	6.88	—	27	< 0.001***	—	21.02	—	27	< 0.001***	—	20.96	—	27	< 0.001***	—
Time	0.08	0.01	166	0.89	—	0.93	0.11	168	0.26	—	0.42	0.05	165	0.41	—
Count training	−0.85	−0.08	27	0.58	—	−0.36	−0.04	27	0.78	—	−0.13	−0.02	27	0.88	—
NL training	−0.79	−0.07	27	0.61	—	−0.23	−0.03	27	0.86	—	0.07	0.01	27	0.95	—
Time × count training	2.04	0.15	166	< 0.01**	0.02	−0.62	−0.05	168	0.58	0.08	0.18	0.02	165	0.84	0.09
Time × NL training	0.78	0.06	166	0.34	0.07	0.17	0.01	168	0.88	0.10	−0.87	−0.08	165	0.17	0.04
Symb. number lines						Non-symb. number lines						Counting (ENT-R)			
Model intercept	0.40	—	27	< 0.001***	—	0.47	—	27	< 0.001***	—	11.64	—	27	< 0.001***	—
Time	−0.01	−0.01	166	0.89	—	0.16	0.32	165	< 0.001***	—	1.61	0.18	168	< 0.01**	—
Count training	−0.08	−0.14	27	0.32	—	0.07	0.13	27	0.33	—	−0.50	−0.01	27	0.74	—
NL training	−0.05	−0.09	27	0.55	—	0.07	0.13	27	0.34	—	−0.27	−0.03	27	0.86	—
Time × count training	0.16	0.23	166	< 0.01**	0.01	−0.05	−0.08	165	0.30	0.06	1.69	0.14	168	0.03*	0.03
Time × NL training	0.10	0.14	166	0.08	0.03	−0.05	−0.08	165	0.31	0.06	0.88	0.07	168	0.25	0.05

The control group served as a reference group in the presented analyses. NL, number line; Non-symb., non-symbolic; B, unstandardized coefficient; β , standardized coefficient; α , adjusted alpha after Benjamini-Hochberg correction. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

which have been hypothesized to advance mapping capacities in young children (Lipton and Spelke, 2005; Le Corre and Carey, 2007; Siegler and Ramani, 2009; Fisher et al., 2011).

Results indicate that kindergartners outperformed the control group only after counting training. This implies that number processing and consequent arithmetic skills can be nurtured through counting activities (Lipton and Spelke, 2005; Le Corre and Carey, 2007), and it may suggest that development in the school context is also furthered by counting more than by number line training. Formation of mapping skills may occur through the repeated bottom-up process of matching number words with visible quantities, as suggested by Le Corre and Carey (2007), and quantities may more easily be processed by assessing individual items in a set one-by-one than by placement in a higher-order framework, which is done in number line tasks using top-down processing. This more fluent processing may have led to significant gains in the counting group and not the number line group in comparison to the control group, even when measuring progress through a number line task. This is congruent with the notion that mapping fosters arithmetic skills by making symbolic numbers more meaningful to children (Wong et al., 2016); something that is likely more easily achieved through a tangible and observable counting process than through an abstract number line game. It should be noted though that differences in gains between the counting group and the number line group were not significant. Rather, the number line group made small (non-significant) progress on several tasks that imply that with sufficient power, number line activities would in fact show small effects on number sense measures, although not in the same order of magnitude as the counting training, nor would they be of the same order as the effects previously reported (e.g., Siegler and Ramani, 2009; Maertens et al., 2016).

Advances in mapping could be seen in the number line task, measuring mapping, but not on the symbolic comparison task, scores on which can also be seen as an indication of mapping skills (Kolkman et al., 2013). This may be due to the low difficulty of the task. Numbers ranged up to 9, and children showed no obvious difficulties completing the task. This is also apparent from their scores at pretest, during which children performed well above the chance level of 50%. Possibly, the few mistakes made by children were due to other factors such as attentional resources, rather than their mapping capacities.

It is also worth noting that the counting training had no effects on measures of non-symbolic processing. Effects of number sense training on non-symbolic tasks have previously been found to be lacking (Malofeeva et al., 2004) or to have smaller effects than on symbolic tasks (Wilson et al., 2006), suggesting that it is primarily the symbolic skill level that interacts with broader numerical development and plays a key-role in the development of number sense. It has been hypothesized that non-symbolic skills serve as a foundation for all further development in mathematics and number skills (Dehaene, 2001), but the current study suggests that limited gains in non-symbolic skills do not constrain gains in symbolic skills.

Younger children made somewhat greater gains during the counting training than older children in arithmetic and number line scores. This may be due to a difference in time spent at

school between the children. The correlation of scores with the age of the children may be indicative of a catch-up effect in younger children, after more instruction. However, the absence of correlations between age and most measures at pretest indicates that this explanation does not sufficiently explain the current results. Alternatively, younger children may have found the activities from the training more appealing, or they may have complied more with instructions set by the trainer, resulting in greater training effects. The effects are contrary to the results presented in the meta-analysis by Kroesbergen and Van Luit (2003), who reported greater training effects of older children. It should be noted, however, that these concerned between-study differences, which may be the result of differences between trainings, and that this is not necessarily indicative of similar within-group moderation effects.

The finding that training gains are moderated by the SES of children (Starkey et al., 2004) could not be replicated. The absence of a moderation effect of SES may be caused by the criterion for group membership. In the current study, children were classified based on the educational level of the parents, while children in the study by Starkey et al. (2004) were classified based on parental income. Although both are indicative of SES, these constructs may have different implications for child development. More specifically, any difference in material resources such as educational materials for children, that may have been associated with differences in training gains in the cited study, may not have been relevant for the groups constructed in the current study. A second cause of the disparity might be the inequality in incomes between families, which is smaller in the Netherlands than in the United States (Central Intelligence Agency, n.d.) and may therefore have smaller consequences for child outcomes.

Future research is needed to elaborate on the parameters of similar training programs. For example, it may be investigated what the effects are of the duration of a training. A meta-analysis concerning the effects of mathematics and number sense trainings has suggested that longer trainings yielded smaller training gains (Kroesbergen and Van Luit, 2003). However, the authors proposed that this was due to differences in scope of the training studies: shorter trainings aimed to improve a more narrow range of skills, leading to more improvement in fewer skills. In a study investigating two training programs with a similar scope, greater training gains and more transfer were reported for the training with the more extensive time span (Toll and Van Luit, 2014). This difference in training gains was significant for general mathematics, and marginally significant for arithmetic. Other evidence concerning the duration of training is scarce, although effects of very short number intervention studies of only four sessions have been reported (Ramani and Siegler, 2008, 2011; Whyte and Bull, 2008).

Also, the range of numbers included is a topic that may be investigated in future research. In the current study, numbers up to 50 were included in the training programs, but other studies have reported on trainings using number ranges up to 10 (Siegler and Ramani, 2008, 2009), up to 15 (Van Luit and Schopman, 2000; Blöte et al., 2006), up to 20 (Fisher et al., 2011) or up to 21 (Baroody et al., 2009). It is likely that children of different ages

benefit to a different extent from training programs that focus on different number ranges, and that older children benefit more from broader number ranges as they are already familiar with smaller numbers. However, the exact effect of the inclusion of different number ranges in training programs is as yet unknown.

A limitation of the current study is the sample selection. In the current study, all children were eligible for participation, while not every child was in direct need of a number sense intervention. This may have limited the gains children made during the trainings compared to the control group: children not at-risk for delays in number sense typically make gains in number sense that are sufficient to start formal education without intervention, explaining gains in the control group. Also, longitudinal studies are needed to map the benefits of the interventions fully. Finally, the matching procedure in the current study, in which children were matched at school-level, ensured great variation in number knowledge between children in each training group. Smaller variation in number knowledge may be more beneficial to training gains, because of a more equal level between children at the start of the training, making activities similarly useful to all children in a training group.

A second limitation is the number range covered by the tasks used to evaluate children's progress in numerical skills. This number range differed per task, with number line tasks ranging up to 100, arithmetic and counting items dealing with quantities up to 20, and comparison tasks only ranging up to 9. This difference in tasks hampers a full comparison in progress between tasks. Conclusions, therefore, can only be made with regard to the comparison in progress between experimental groups and the control group, and not with regard to any difference in progress between various tasks used to index numerical skills. Moreover, number ranges covered during the training sessions only partially overlapped with the pre- and post-tests. Perhaps training gains would be larger if the same number ranges were covered in the training tasks.

Nevertheless, the current study adds to the body of literature by providing experimental evidence for the importance of

counting to advance mapping skills and arithmetic skills, and the smaller, non-significant training gains after a number line training. Non-symbolic skills were not influenced by training at all. These findings are of both theoretical and practical significance, because of the implications they have for theories concerning the building of mapping skills and its consequence for arithmetic development, and because of the clear distinction they make in effectiveness of different training activities, which has clear and large implications for the effectiveness of school curricula focusing on number sense.

AUTHOR CONTRIBUTIONS

IF-vdB was in charge of the data collection, supervising undergraduate research assistants, and designing the materials, as well as writing the main part of the manuscript. EK checked the data collection decisions and materials, gave advice to improve the materials, and contributed to the manuscript with advice about the theoretical framing of the study as well as suggestions with regard to presentation, and by adding improvements to the text. She wrote major parts of the research proposal based on which the study was funded. JVL contributed to the manuscript with advice about the theoretical framing of the study as well as suggestions with regard to presentation and by adding improvements to the text. He wrote major parts of the research proposal based on which the study was funded.

FUNDING

This study was supported by a project grant from the Netherlands Organisation for Scientific Research (NWO), grant 411-07-113.

ACKNOWLEDGMENTS

We would like to thank the participating children, their parents, and teachers for their participation and consent.

REFERENCES

- Association for Universities in the Netherlands (2004). *The Netherlands Code of Conduct for Academic Practice: Principles of Good Academic Teaching and Research*. Available at: <https://www.uu.nl/en/files/the-netherlands-code-of-conduct-for-academic-practice-2004-version-2014pdf>
- Baroody, A. J., Eiland, M., and Thompson, B. (2009). Fostering at-risk preschoolers' number sense. *Early Educ. Dev.* 20, 80–128. doi: 10.1080/10409280802206619
- Barth, H. C., and Paladino, A. M. (2011). The development of numerical estimation: evidence against a representational shift. *Dev. Sci.* 14, 125–135. doi: 10.1111/j.1467-7687.2010.00962.x
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57, 289–300. doi: 10.2307/2346101
- Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., and Zorzi, M. (2010). Numerical estimation in preschoolers. *Dev. Psychol.* 46, 545–551. doi: 10.1037/a0017887
- Blöte, A. W., Liefvering, L. M., and Ouweland, K. (2006). The development of many-to-one counting in 4-year-old children. *Cogn. Dev.* 21, 332–348. doi: 10.1016/j.cogdev.2005.10.002
- Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 41, 189–201. doi: 10.1037/0012-1649.41.6.189
- Booth, J. L., and Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Dev.* 79, 1016–1031. doi: 10.1111/j.1467-8624.2008.01173.x
- Central Intelligence Agency (n.d.). *The World Factbook: Distribution of Family Income – Gini Index (ISSN 1553-8133)*. Available at: <https://www.cia.gov/library/publications/the-world-factbook/fields/2172.html>
- Clarke, B., and Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychol. Rev.* 33, 234–248.
- Dackermann, T., Fischer, U., Huber, S., Nuerk, H.-C., and Moeller, K. (2016). Training the equidistant principle of number line spacing. *Cogn. Process.* 17, 243–258. doi: 10.1007/s10339-016-0763-8

- Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. New York, NY: Oxford University Press.
- Dehaene, S. (2001). Précis of the number sense. *Mind Lang.* 16, 16–36. doi: 10.1111/1468-0017.00154
- Desoete, A., Ceulemans, A., De Weerd, F., and Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? Findings from a longitudinal study. *Br. J. Educ. Psychol.* 82, 64–81. doi: 10.1348/2044-8279.002002
- Desoete, A., Praet, M., Titeca, D., and Ceulemans, A. (2013). Cognitive phenotype of mathematical learning disabilities: what can we learn from siblings? *Res. Dev. Disabil.* 34, 404–412. doi: 10.1016/j.ridd.2012.08.022
- Dyson, N. I., Jordan, N. C., and Glutting, J. (2013). A number sense intervention for low-income kindergartners at risk for mathematics difficulties. *J. Learn. Disabil.* 46, 166–181. doi: 10.1177/0022219411410233
- Fisher, U., Moeller, K., Bientzle, M., Cress, U., and Nuerk, H.-C. (2011). Sensorimotor spatial training of number magnitude representation. *Psychon. Bull. Rev.* 18, 177–183. doi: 10.3758/s13423-010-0031-3
- Friso-van den Bos, I., Kolkman, M. E., Kroesbergen, E. H., and Leseman, P. P. M. (2014). Explaining variability: numerical representations in 4- to 8-year old children. *J. Cogn. Dev.* 33, 23–29. doi: 10.1080/15248372.2012.742900
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., and Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Dev.* 78, 1343–1359. doi: 10.1111/j.1467-8624.2007.01069.x
- Gersten, R., and Chard, D. J. (1999). Number sense: rethinking arithmetic instruction for students with mathematical disabilities. *J. Spec. Educ.* 33, 18–28. doi: 10.1177/002246699903300102
- Gracia-Bafalluy, M., and Noël, M.-P. (2008). Does finger training increase young children's numerical performance? *Cortex* 44, 368–375. doi: 10.1016/j.cortex.2007.08.020
- Hornung, C., Schiltz, C., Brunner, M., and Martin, R. (2014). Predicting first-grade mathematics achievement: the contributions of domain-general cognitive abilities, nonverbal number sense, and early number competence. *Front. Psychol.* 5:272. doi: 10.3389/fpsyg.2014.00272
- Jordan, N. C., Glutting, J., Dyson, N. I., Hassinger-Das, B., and Irwin, C. (2012). Building kindergartners' number sense: a randomized controlled study. *J. Exp. Child Psychol.* 104, 647–660. doi: 10.1037/a0029018
- Kolkman, M. E., Kroesbergen, E. H., and Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learn. Instr.* 25, 95–103. doi: 10.1016/j.learninstruc.2012.12.001
- Krajewski, K., Nieding, G., and Schneider, W. (2008). Kurz- und langfristige Effekte mathematischer Frühförderung im Kindergarten durch das Programm "Mengen, zählen, Zahlen" [Short- and long-term effects of early mathematical support in kindergarten through the programme "Quantities, counting, numbers"]. *J. Entwicklungspsychol. Pädagog. Psychol.* 40, 135–146. doi: 10.1026/0049-8637.40.3.135
- Kroesbergen, E. H., and Van Luit, J. E. H. (2003). Mathematics interventions for children with special educational needs. *Remedial Spec. Educ.* 24, 97–114. doi: 10.1177/07419325030240020501
- Kroesbergen, E. H., Van 't Noordende, J. E., and Kolkman, M. E. (2012). "Number sense in low-performing kindergarten children: effects of a working memory and a number sense training" in *Reading, Writing, Mathematics and the Developing Brain: Listening to Many Voices*, eds Z. Breznitz, O. Rubinsten, V. J. Molfese, and D. Molfese (New York, NY: Springer Publications), 295–313. doi: 10.1007/978-94-007-4086-0_16
- Laski, E. V., and Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Dev.* 78, 1723–1743. doi: 10.1111/j.1467-8624.2007.01087.x
- Le Corre, M., and Carey, S. (2007). One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition* 105, 395–438. doi: 10.1016/j.cognition.2006.10.005
- Lipton, J. S., and Spelke, E. S. (2005). Preschool children's mapping of number words to nonsymbolic numerosities. *Child Dev.* 76, 978–988. doi: 10.1111/j.1467-8624.2005.00891.x
- Maertens, B., De Smedt, B., Sasanguie, D., Elen, J., and Reynvoet, B. (2016). Enhancing arithmetic in pre-schoolers with comparison or number line estimation training: does it matter? *Learn. Instr.* 46, 1–11. doi: 10.1016/j.learninstruc.2016.08.004
- Malofeeva, E., Day, J., Saco, X., Young, L., and Ciancio, D. (2004). Construction and evaluation of a number sense test with head start children. *J. Educ. Psychol.* 96, 648–659. doi: 10.1037/0022-0663.96.4.648
- Moeller, K., Martignon, L., Wessolowski, S., Engel, J., and Nuerk, H.-C. (2011). Effects of finger counting on numerical development – The opposing views of neurocognition and mathematics education. *Front. Psychol. Cogn.* 2:328. doi: 10.3389/fpsyg.2011.00328
- Mundy, E., and Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *J. Exp. Child Psychol.* 103, 490–502. doi: 10.1016/j.jecp.2009.02.003
- Noël, M.-P. (2005). Finger gnosis: a predictor of numerical abilities in children? *Child Neuropsychol.* 11, 413–430. doi: 10.1080/09297040590951550
- Noël, M.-P., and Rousselle, L. (2011). Developmental changes in the profiles of dyscalculia: an explanation based on a double exact-and-approximate number representation model. *Front. Hum. Neurosci.* 5:165. doi: 10.3389/fnhum.2011.00165
- Passolunghi, M.-C., and Lanfranchi, S. (2012). Domain-specific and domain-general precursors of mathematical achievement: a longitudinal study from kindergarten to first grade. *Br. J. Educ. Psychol.* 82, 42–63. doi: 10.1111/j.2044-8279.2011.02039.x
- Piaget, J. (1970). *Genetic Epistemology*. New York, NY: Norton.
- Ramani, G. B., and Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Dev.* 79, 375–394. doi: 10.1111/j.1467-8624.2007.01131.x
- Ramani, G. B., and Siegler, R. S. (2011). Reducing the gap in numerical knowledge between low- and middle-income preschoolers. *J. Appl. Dev. Psychol.* 32, 146–159. doi: 10.1016/j.appdev.2011.02.005
- Schneider, M., Merz, S., Stricker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., et al. (2018). Associations of number line estimation with mathematical competence: a meta-analysis. *Child Dev.* doi: 10.1111/cdev.13068 [Epub ahead of print].
- Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi: 10.1111/j.1467-8624.2004.00684.x
- Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–244. doi: 10.1016/j.cogpsych.2004.01.002
- Siegler, R. S., and Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Dev. Sci.* 11, 655–661. doi: 10.1111/j.1467-7687.2008.00714.x
- Siegler, R. S., and Ramani, G. B. (2009). Playing linear number board games – but not circular ones – improves low-income preschoolers' numerical understanding. *J. Educ. Psychol.* 101, 545–560. doi: 10.1037/a0014239
- Simms, V., Muldoon, K., and Towse, J. (2013). Plane thinking: mental representations in number line estimation as a function of orientation, scale, and counting proficiency. *J. Exp. Child Psychol.* 115, 468–480. doi: 10.1016/j.jecp.2013.03.011
- Spelke, E. S. (2000). Core knowledge. *Am. Psychol.* 55, 1233–1243. doi: 10.1037/0003-066X.55.11.1233
- Starkey, P., Klein, A., and Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Child. Res. Q.* 19, 99–120. doi: 10.1016/j.ecresq.2004.01.002
- Toll, S. W. M., and Van Luit, J. E. H. (2014). Effects of remedial numeracy instruction throughout kindergarten starting at different ages: evidence from a large-scale longitudinal study. *Learn. Instr.* 33, 39–49. doi: 10.1016/j.learninstruc.2014.03.003
- Van Luit, J. E. H., and Schopman, E. A. M. (2000). Improving early numeracy of young children with special educational needs. *Remedial Spec. Educ.* 21, 27–40. doi: 10.1177/074193250002100105
- Van Luit, J. E. H., and Van de Rijt, B. A. M. (2009). *Utrechtse Getalbegrip Toets-Revised [Early Numeracy Test-Revised]*. Doetinchem: Graviant.
- Whyte, J. C., and Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschooler. *Dev. Psychol.* 44, 588–596. doi: 10.1037/0012-1649.44.2.588

- Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., and Dehaene, S. (2006). An open trial assessment of "The Number Race", An adaptive computer game for remediation of dyscalculia. *Behav. Brain Funct.* 2:20. doi: 10.1186/1744-9081-2-20
- Wong, T. T.-Y., Ho, C. S. H., and Tang, J. (2016). The relation between ANS and symbolic arithmetic skills: the mediating role of number-numerosity mappings. *Contemp. Educ. Psychol.* 46, 208–217. doi: 10.1016/j.cedpsych.2016.06.003
- Zhang, D., Ding, Y., Stegall, J., and Mo, L. (2012). The effect of visual-chunking-representation accommodation on geometry testing for students with math disabilities. *Learn. Disabil. Res. Pract.* 27, 167–177. doi: 10.1111/j.1540-5826.2012.00364.x

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

Copyright © 2018 Friso-van den Bos, Kroesbergen and Van Luit. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Different Subcomponents of Executive Functioning Predict Different Growth Parameters in Mathematics: Evidence From a 4-Year Longitudinal Study With Chinese Children

Wei Wei¹, Liyue Guo², George K. Georgiou³, Athanasios Tavouktsoglou⁴ and Ciping Deng^{2*}

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

Laura Visu-Petra,
Babeş-Bolyai University, Romania
Bert De Smedt,
KU Leuven, Belgium
Julie Ann Jordan,
Queen's University Belfast,
United Kingdom

*Correspondence:

Ciping Deng
cpdeng@psy.ecnu.edu.cn

Specialty section:

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

Received: 25 January 2018

Accepted: 01 June 2018

Published: 21 June 2018

Citation:

Wei W, Guo L, Georgiou GK,
Tavouktsoglou A and Deng C (2018)
Different Subcomponents
of Executive Functioning Predict
Different Growth Parameters
in Mathematics: Evidence From
a 4-Year Longitudinal Study With
Chinese Children.
Front. Psychol. 9:1037.
doi: 10.3389/fpsyg.2018.01037

¹ College of Education, Shanghai Normal University, Shanghai, China, ² Shanghai Key Laboratory of Brain Functional Genomics, Changning-ECNU Mental Health Center, School of Psychology and Cognitive Science, East China Normal University, Shanghai, China, ³ Department of Educational Psychology, University of Alberta, Edmonton, AB, Canada, ⁴ Faculty of Science, Concordia University of Edmonton, Edmonton, AB, Canada

Executive functioning (EF), an umbrella term used to represent cognitive skills engaged in goal-directed behaviors, has been found to be a unique predictor of mathematics performance. However, very few studies have examined how the three core EF subcomponents (inhibition, shifting, and working memory) predict the growth parameters (intercept and slope) in mathematics skills and even fewer studies have been conducted in a non-Western country. Thus, the purpose of this study was to examine how inhibition, shifting, and working memory predict the growth parameters in arithmetic accuracy and fluency in a group of Chinese children ($n = 179$) followed from Grade 2 (mean age = 97.89 months) to Grade 5 (mean age = 133.43 months). In Grade 2, children were assessed on measures of nonverbal IQ, number sense, speed of processing, inhibition, shifting, and working memory. In addition, in Grades 2–5, they were assessed on arithmetic accuracy and fluency. Results of structural equation modeling showed that nonverbal IQ, speed of processing, and number sense predicted the intercept in arithmetic accuracy, while working memory was the only EF subcomponent to predict the slope (rate of growth) in arithmetic accuracy. In turn, number sense, speed of processing, inhibition, and shifting were significant predictors of the intercept in arithmetic fluency. None of the EF subcomponents predicted the slope in arithmetic fluency. Our findings reinforce those of previous studies in North America and Europe showing that EF contributes to mathematics performance over and above other key predictors of mathematics, and suggest that different EF subcomponents may predict different growth parameters in mathematics.

Keywords: executive functioning, working memory, arithmetic, mathematics, Chinese, longitudinal

INTRODUCTION

Executive functioning (EF), an umbrella term used to represent cognitive skills engaged in goal-directed behaviors, such as inhibition, mental flexibility, and working memory (e.g., Chan et al., 2008; Best et al., 2009; Diamond, 2013), is important not only for behavioral regulation in classroom that ultimately enhances learning (e.g., Day et al., 2015), but also for the development of specific cognitive skills that further support children's academic performance (e.g., Fuhs et al., 2016). One of the academic skills that EF appears to make a unique contribution to is mathematics (e.g., Espy et al., 2004; Blair and Razza, 2007; Willoughby et al., 2012; Allan et al., 2014). Despite the acknowledged importance of EF in mathematics performance, far less is known about how EF subcomponents predict the growth parameters (intercept and slope) of mathematics development. Therefore, the purpose of this study was to examine how the three core EF subcomponents (inhibition, shifting, and working memory) predict the growth parameters of two different mathematics skills (arithmetic accuracy and fluency) in a sample of Chinese children followed from Grade 2–5.

Executive functioning has been conceptualized as a multicomponent construct composed of inhibition, shifting, and working memory (e.g., Miyake et al., 2000; Lehto et al., 2003; Wu et al., 2011; Xu et al., 2013; however, see also Testa et al., 2012, for more EF subcomponents). Inhibition, defined as the ability of an individual to override a dominant but inappropriate response, may help children suppress inappropriate strategies while operating on math problems or suppress the prepotent activation of an inappropriate number representation (Bull and Lee, 2014). In turn, shifting, defined as flexibly switching attention between different mindsets, may help individuals switch between different operation rules. Finally, working memory, defined as the ability of an individual to hold information in short-term memory (storage) while processing some other (processing), is needed when solving different mathematics problems [e.g., $(3 + 5) * 4 = ?$] because individuals need to first hold part of the solution in their memory (e.g., the result of $3 + 5$) before executing another operation (e.g., multiplying by 4).

Among the three EF subcomponents, working memory is perhaps the most studied in relation to mathematics (see Raghubar et al., 2010, for a review). Two meta-analyses have reported a moderate correlation between working memory and different mathematics skills (r s ranged from 0.31 to 0.38 in Friso-van den Bos et al., 2013; the average correlation was 0.35 in Peng et al., 2015). Recent studies have also shown that inhibition and shifting are significant correlates of mathematics performance (e.g., Blair and Razza, 2007; Andersson, 2008; Clark et al., 2010; Willoughby et al., 2012; Gilmore et al., 2015; Cragg et al., 2017; Purpura et al., 2017), although the strength of their relationship appears to be lower compared to that of working memory. In their meta-analyses, Allan et al. (2014) and Friso-van den Bos et al. (2013) reported an average correlation of 0.27 between inhibition and mathematics. Likewise, Yeniad et al. (2013) and Friso-van den Bos et al. (2013) found the correlation between shifting and mathematics to be 0.26 and 0.28, respectively.

Unfortunately, most previous EF studies have focused on the role of individual EF subcomponents and, as a result, we do not know how they predict mathematics skills in the presence of each other. In addition, the few studies that have included all three EF subcomponents have produced mixed findings (e.g., Bull and Scerif, 2001; Andersson, 2008; Agostino et al., 2010; van der Ven et al., 2012; Cantin et al., 2016; Lubin et al., 2016; Cragg et al., 2017; Vandembroucke et al., 2017). For example, whereas some studies have found working memory to account for unique variance in mathematics skills after controlling for the effects of all other EF subcomponents (e.g., Bull and Scerif, 2001; Agostino et al., 2010; Lubin et al., 2016; Cragg et al., 2017; Vandembroucke et al., 2017), others failed to find any significant effects (e.g., Espy et al., 2004; Cantin et al., 2016). Similarly, whereas some studies have found that inhibition and shifting make a unique contribution to mathematics skills (e.g., Espy et al., 2004; Andersson, 2008; Cantin et al., 2016; Cragg et al., 2017; Purpura et al., 2017), others did not (e.g., Monette et al., 2011; Rose et al., 2011; Lee et al., 2012; Vandembroucke et al., 2017).

There might be two reasons for the mixed findings. First, they may reflect differential effects of EF subcomponents on different mathematics skills. Mathematics skills consist of several components themselves including arithmetic accuracy (the accuracy of performing different operations either by using procedural or retrieval strategies) and arithmetic fluency (the speed with which different arithmetic problems are solved). In studies in which math accuracy scores were used, working memory was found to make a unique contribution (e.g., Andersson, 2008; Agostino et al., 2010; Lee et al., 2012; Cragg et al., 2017). In contrast, in studies in which fluency scores were used, working memory did not predict mathematics (e.g., Andersson, 2008; Cantin et al., 2016; Purpura et al., 2017). The opposite pattern appears to be true for inhibition and shifting. Studies have reported a unique contribution of shifting and inhibition to arithmetic fluency (e.g., Andersson, 2008; Clark et al., 2010; Cragg et al., 2017) but not to accuracy (e.g., Agostino et al., 2010; Lee et al., 2012). Cragg and Gilmore (2014) concluded that the contribution of EF subcomponents may differ across different aspects of mathematics skills.

Second, most previous studies examining the contribution of inhibition and shifting to mathematics skills have administered speeded measures of both, without controlling for the effects of speed of processing. As van der Sluis et al. (2007), and more recently Georgiou and Das (2018) have indicated, in this kind of studies unless researchers control for speed of processing we do not know if the effects of EF on mathematics are driven by their executive processing demands or by speed. Most of the EF tasks, especially the inhibition and shifting tasks, are speeded because of ceiling effects in accuracy (e.g., Anderson, 2002; Lee et al., 2012). The results of a meta-analysis by Yeniad et al. (2013) showed that the average correlation between response time measures of shifting and mathematics ($r = 0.36$) was higher than that between accuracy measures of shifting and mathematics ($r = 0.25$). Rose et al. (2011) and Bull and Lee (2014) further argued that the variance in mathematics skills explained by EF may be attributed to speed of processing, because speed of processing, as a domain-general cognitive skill, also contributes to mathematics. Thus, the

contribution of EF subcomponents, especially of inhibition and shifting, may decrease when the effects of speed of processing are controlled. In line with this prediction, some studies have found that inhibition was no longer predicting the mathematics skills when the effects of speed of processing were controlled (Rose et al., 2011; Purpura et al., 2017). Fuchs et al. (2006) also showed that working memory did not explain any unique variance in mathematics skills after controlling for the effects of speed of processing and nonverbal IQ. Certainly, these findings need to be replicated.

Beyond the contradictory findings of previous studies that included all three EF subcomponents, previous studies examining the role of EF in mathematics skills suffer from at least three limitations. First, most previous studies have not examined the role of the EF subcomponents in the presence of other key predictors of mathematics such as number sense. Number sense refers to an individual's "fluidity and flexibility with numbers," which includes skills such as understanding what numbers mean and how they relate to each other (Gersten and Chard, 1999). The first reason why the effects of number sense should be partialled out is that some EF tasks (e.g., Trail Making) typically use numbers as their stimuli and this may inflate the relations with mathematics (Cragg and Gilmore, 2014). In addition, although some previous studies have shown that earlier EF predicts future number competence (e.g., Kolkman et al., 2013; McClelland et al., 2014; Purpura et al., 2017), little is known about whether EF continues to predict mathematics skills after controlling for number competence such as number sense. Fuhs et al. (2016), for example, found that the effects of early EF on concurrent mathematics performance were fully mediated by number sense, and Simanowski and Krajewski (2017) also found that EF in kindergarten could not predict mathematics skills in Grades 1 and 2 (mean ages were 87 and 99 months, respectively) after controlling for early number competence. Therefore, as Viterbori et al. (2015) have suggested, children's number sense should be controlled before examining the contribution of EF subcomponents to mathematics skills.

Second, most previous studies examining the relationship between EF and mathematics are cross-sectional (e.g., Agostino et al., 2010; Rose et al., 2011; Cantin et al., 2016). The few longitudinal studies (e.g., Swanson, 2006; McClelland et al., 2014; Simanowski and Krajewski, 2017; Vandenbroucke et al., 2017) have covered only a limited developmental span (most often from Kindergarten to Grades 1 and 2) and have used the EF skills (assessed at an earlier point in time) to predict mathematics skills at a later point in time (often assessed once). To our knowledge, only a handful of longitudinal studies have examined how EF predicts different growth parameters (intercept and slope) in mathematics (see Bull et al., 2008; Geary, 2011; van der Ven et al., 2012; Van de Weijer-Bergsma et al., 2015; Lee and Bull, 2016), and of these studies only two had assessed all three EF subcomponents (Bull et al., 2008; van der Ven et al., 2012). The results of van der Ven et al. (2012) showed that working memory (updating) in Grade 1 (mean age = 77 months) correlated with the intercept in mathematics (a comprehensive mathematics test) during Grades 1 and 2 (mean age = 95 months), while a factor composed of inhibition and shifting in Grade 1 did not correlate

with either growth parameter. Similarly, the results of Bull et al. (2008) showed that working memory along with inhibition at kindergarten (mean age = 54 months) predicted the intercept in mathematics during Grade 1 (5–6 years old) and Grade 3 (7–8 years old). Furthermore, Geary (2011) and Lee and Bull (2016) examined the growth parameter of arithmetic accuracy (assessed with numerical operations) during a longer span (more than 3 years), and found working memory also predicted the slope in arithmetic accuracy. Another study, Van de Weijer-Bergsma et al. (2015), found working memory at the beginning of Grade 2 (6–8 years old) correlated with the intercept not the slope in math fluency during Grade 2. Thus, more research is needed on how all three EF subcomponents predict the growth parameters of mathematics development.

Finally, almost all of the studies reviewed above were conducted in North America or Europe and we do not know if their findings generalize to East Asian countries (e.g., China). There are reasons to believe that the role of EF subcomponents may be different in China than in Western countries. The first reason relates to the role of working memory. Because Chinese digits are monosyllabic and have a shorter pronunciation duration they allow individuals to hold a larger number of digits in their short-term memory. If simple calculations can be solved with direct retrieval of facts from long-term memory, then individuals with a larger pool of arithmetic facts in their memory should also have superior performance in calculations. Indeed, a few cross-cultural studies have shown that Chinese outperform North Americans in mental calculation (e.g., Stevenson et al., 1990; Campbell and Xue, 2001; Wang and Lin, 2009; Lonnemann et al., 2016). Imbo and LeFevre (2009) also showed that Chinese university students required fewer working memory resources than Belgian or Canadian university students when solving complex addition problems. If Chinese children solve simple addition and subtraction problems by relying on rote memorization, then the contribution of working memory may not be as strong as it has been reported in previous studies in North America. Some studies have provided evidence in support of this hypothesis (e.g., Geary et al., 1996; Thorell et al., 2013; Cui et al., 2017), but more research is needed.

Second, inhibition may be less important for mathematics skills among Chinese children than among North-American children. Geary et al. (1996) found that more than half of American children in Grade 2 (mean age = 94 months) and Grade 3 (mean age = 104 months) were still using basic strategies in addition, such as counting fingers and verbal counting, while almost all Chinese children in the same grades (with comparable mean ages) were relying on direct retrieval. Relying on strategies such as verbal counting may lead to one of the most common errors in calculation, i.e., a counting-string associate of one of the addends (e.g., $3 + 5 = 6$; Siegler and Shrager, 1984). To avoid this error, American children have to actively suppress any irrelevant association when retrieving arithmetic facts from long-term memory. In contrast, Chinese children may not need to inhibit irrelevant associations if they directly retrieve the answers to calculations from their long-term memory. Indeed, Lan et al. (2011) found that inhibition of Chinese preschoolers uniquely predicted counting, but failed to predict calculation,

while inhibition of American children uniquely predicted both counting and calculation. Similarly, Peng et al. (2012) found that performance on a color-word Stroop task (one of the most widely used measures of inhibition) failed to differentiate between Chinese fifth-graders with mathematics difficulties and their typically developing peers (the mean age of both groups was 132 months).

To our knowledge, only four studies have examined the contribution of EF to mathematics skills among Chinese children and all of them have focused on the concurrent relationships between some EF subcomponents and mathematics during kindergarten. Three studies with Chinese preschoolers (Zhang, 2016; Chung et al., 2017; Zhang et al., 2017) showed that inhibition and working memory or an EF factor composed of inhibition and working memory uniquely predicted early mathematics skills after controlling for rapid naming, vocabulary, and visual skills. Another study (Lan et al., 2011) found that Chinese preschoolers' inhibition predicted counting, but failed to predict calculation. Working memory predicted both counting and calculation. Therefore, it remains unclear whether EF subcomponents can predict mathematics skills longitudinally, especially the growth rate of mathematics skills.

The Present Study

The purpose of this study was to examine how the three core EF subcomponents (inhibition, shifting, and working memory) predict the growth parameters (intercept and slope) of arithmetic accuracy and fluency in a group of Chinese children followed from Grade 2 to 5. Based on the findings of previous studies that examined the predictors of growth parameters in mathematics performance (see Bull et al., 2008; Geary, 2011; van der Ven et al., 2012; Van de Weijer-Bergsma et al., 2015; Lee and Bull, 2016), we expected that:

- (1) Working memory would predict both growth parameters of arithmetic accuracy (see Geary, 2011; Lee and Bull, 2016), and the intercept of arithmetic fluency (see Van de Weijer-Bergsma et al., 2015),
- (2) Inhibition would predict only the intercept of arithmetic fluency (see Bull et al., 2008) and,
- (3) Shifting would not predict any growth parameter in any mathematics skill (see van der Ven et al., 2012).

MATERIALS AND METHODS

Participants

One hundred seventy-nine Grade 2 Chinese children (82 girls and 97 boys; mean age = 97.89 months, $SD = 3.56$) were recruited on a voluntary basis from public schools in Shanghai (China) to participate in the study (T1). The children were reassessed in Grades 3, 4, and 5 (T2, T3, and T4), when they were 109.65 ($SD = 3.62$), 122.99 ($SD = 3.55$) and 133.43 ($SD = 3.70$) months old, respectively. By Grade 5, only 165 children (or 92% of the original sample) remained in the study. The children who withdrew from the study did not differ significantly from the children who remained in the study on any of the measures

administered in Grade 2 (all $ps > 0.10$). All children were native speakers of Mandarin and none was experiencing any intellectual, sensory, or behavioral difficulties (based on teachers' reports). Most of the children came from families of middle socioeconomic background (based on parents' occupation and education). Parental permission and ethical approval from the Research Ethics Committee of East China Normal University was obtained prior to testing.

Materials

Nonverbal IQ

To assess nonverbal IQ we administered the Nonverbal Matrices task from the Das–Naglieri Cognitive Assessment System (DN CAS) battery (Naglieri and Das, 1997). This task has been used in several previous studies in Chinese showing good reliability and validity evidence (e.g., Liao et al., 2008; Deng et al., 2011). Children were presented with a page containing a pattern of shapes/geometric designs that was missing a piece and were asked to choose among five or six alternatives the piece that would accurately complete the pattern. The task was discontinued after four consecutive errors and a participant's score was the total number correct. The Cronbach's alpha reliability coefficient in our sample was 0.94.

Speed of Processing

To assess speed of processing we administered Visual Matching from the Woodcock–Johnson Tests of Cognitive Abilities (Woodcock and Johnson, 1989). Children were presented with 60 rows of numbers and were asked to cross out the two identical numbers in each row (e.g., 8, 9, 5, 2, 9, and 7) within a 3 min time limit. The first 20 rows used single-digit numbers, followed by 20 rows of two-digit numbers, and 20 rows of three-digit numbers. A participant's score was the total number of correctly completed rows. The Cronbach's alpha reliability coefficient in our sample was 0.84.

Number Sense

Number Sets was adopted from Geary et al. (2009) to assess number sense. Children were presented with four pages and each page included a target number at the top of each page (e.g., 5) and sets indicated by two or three linked boxes with Arabic numerals (e.g., 2) and concrete objects (e.g., ●●●). Children were asked to circle all the sets that can be put together to match the target number. The target number of the first two pages was 5 and the time limit was 60 s per page. The target number of the last two pages was 9 and the time limit was 90 s per page. Signal detection method was used to calculate each child's sensitivity (d') in detecting the correct sets based on the number of hits and the number of false alarms (see Geary et al., 2009, for details). The Cronbach's alpha reliability coefficient in our sample was 0.88.

Executive Functioning

Shifting

Shifting was assessed with the Planned Connections task from the DN CAS battery (Naglieri and Das, 1997). Planned Connections is a transparent adaptation of the Trail Making task (Reitan and Wolfson, 1992). In this task, children were presented with

two pages of numbers (1–14) and letters (A–N), and, in each page, they were asked to connect the numbers to the letters in successive order (1, A, 2, B, 3, C, etc.) as fast as possible. The score was the total time to finish both pages. The Cronbach's alpha reliability coefficient in our sample was 0.80.

Inhibition

Inhibition was assessed with the Expressive Attention task from the DN CAS battery (Naglieri and Das, 1997). Expressive Attention is a transparent adaptation of the color-word Stroop task. Children were presented with one page of color rectangles and two pages of Chinese color characters (e.g., 蓝[blue], 黄[yellow], 红[red], 绿[green]). In each page, the stimuli were semi-randomly arranged in eight rows of five. Children were asked to read aloud the color of rectangles in the first page and to name the color characters in the second page as fast as possible. In the third page, children were asked to name as fast as possible the color of the ink in which the color characters were printed (e.g., the character 红[Red] may appear in green ink) instead of saying the color character. A practice page was presented before each trial to ensure all children understood the instructions. The children's response time on the third page was used as a measure of inhibition. The Cronbach's alpha reliability coefficient in our sample was 0.88.

Working memory

The Backward Digit Span task from Wechsler Intelligence Scale for Children-Revised (Wechsler, 1974) was used to assess working memory. In this task, children were asked to repeat a sequence of digits in the reverse order. The strings of digits were presented orally by the experimenter with a time interval of about 1 s between each digit. The strings started with only two digits and one digit was added at each difficulty level (the maximum length was eight digits). The task was discontinued when participants failed both trials of a given length. A participant's score was the maximum length of digit string recalled correctly. The Cronbach's alpha reliability coefficient in our sample was 0.80.

Arithmetic Skills

Arithmetic accuracy

The Numerical Operations task from Wechsler Individual Achievement Test (Wechsler, 2002) was used to assess arithmetic accuracy. There were 61 problems arranged in increasing difficulty that measure arithmetic skills in basic operations (addition, subtraction, multiplication, and division) with integers and fraction, algebra, and geometry. Children were asked to write down the answer to each problem in untimed conditions. A discontinuation rule of four consecutive errors was applied and a child's score was the total number correct. The Cronbach's alpha reliability coefficient in our sample ranged from 0.90 to 0.94.

Arithmetic fluency

To assess arithmetic fluency we administered the Basic Arithmetic Test (BAT, Aunio and Räsänen, 2007, Unpublished). Children were asked to write down the answers to 28 calculation problems within a 3 min time limit. The task consisted of 28 problems: 14 additions (e.g., $2 + 1 = ?$ and $3 + 4 + 6 = ?$) and 14 subtractions (e.g., $4 - 1 = ?$ and $20 - 2 - 4 = ?$) that were mixed

up and presented in two pages. The score was the total number correct divided by the time (in minute) to complete all items. The Cronbach's alpha reliability coefficient in our sample ranged from 0.80 to 0.86.

Procedure

All children were individually assessed in a quiet room at school by the first author and trained graduate students. Testing at all measurement points was completed in April/May (8–9 months after the beginning of the school year). The first testing was completed in two sessions of 30 min each. In Session A, Nonverbal Matrices, Visual Matching, Planned Connections, Expressive Attention, and Backward Digit Span were administered. In Session B, Number Sets, Numerical Operations, and BAT were administered. The order of the tasks within each session was fixed. From T2 to T4, only Numerical Operations and BAT were administered.

Data Analysis

All measures were initially scrutinized for normality. One-way repeated-measures analysis of variance for each arithmetic skill was conducted to examine the main effects of time (linear terms) and time squared (quadratic terms). Pearson correlation coefficients were computed among all variables. Latent growth models were constructed with AMOS 17.0 to predict the growth parameters in each arithmetic skill from the six predictor variables measured. Full information maximum likelihood method was applied to make full use of the data.

RESULTS

Preliminary Data Analysis

Descriptive statistics for all the measures used in our study are shown in **Table 1**. An examination of the distributional properties of the measures revealed that they were within acceptable levels (Tabachnick and Fidell, 2007). The results of one-way repeated-measures analysis of variance for each mathematics skill showed

TABLE 1 | Descriptive statistics for all variables used in the present study.

	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>
Nonverbal IQ	19.77	4.15	12	32
Speed of processing	38.47	5.15	24	51
Number sense	3.14	0.46	1.80	3.83
Inhibition	68	17.25	31	130
Shifting	136.50	46.67	52	292
Working memory	4.13	1.26	2	8
Arithmetic accuracy T1	23.08	2.18	18	31
Arithmetic accuracy T2	28.84	3.14	21	36
Arithmetic accuracy T3	33.49	4.91	25	48
Arithmetic accuracy T4	38.70	7.02	26	54
Arithmetic fluency T1	12.96	3.04	6.96	22.82
Arithmetic fluency T2	15.44	3.90	7.73	31.26
Arithmetic fluency T3	17.88	4.45	8.29	32.36
Arithmetic fluency T4	19.89	4.65	8.67	35.25

a significant main effect of linear terms of time (for arithmetic accuracy, $F(1,163) = 951.19, p < 0.001$; for arithmetic fluency, $F(1,162) = 538.85, p < 0.001$), and a non-significant main effect of quadratic terms of time (for arithmetic accuracy, $F(1,163) = 1.91, p > 0.05$; for arithmetic fluency, $F(1,163) = 2.16, p > 0.05$), which indicated a linear growth trend for both mathematics skills.

Correlations Among All the Measures

Table 2 shows the results of the correlational analysis. There was moderate to high stability between all measurement points for arithmetic accuracy (the correlations ranged from 0.35 to 0.61), and high stability between all measurement points for arithmetic fluency (the correlations ranged from 0.54 to 0.67). Besides, arithmetic accuracy correlated significantly with arithmetic fluency at all measurement points. Nonverbal IQ, speed of processing, number sense, inhibition, and working memory at T1 correlated significantly with arithmetic accuracy at all measurement points (absolute r s values ranged from 0.15 to 0.36), and shifting correlated significantly with arithmetic accuracy at T3 and T4. Speed of processing, number sense, and inhibition at T1 correlated moderately with arithmetic fluency at all measurement points (absolute r s values ranged from 0.30 to 0.43). Finally, working memory at T1 correlated weakly with arithmetic fluency at T2 and T4, and shifting correlated weakly with arithmetic fluency at T3.

Latent Growth Models for Arithmetic Skills

First, unconditional latent linear growth models (without any predictors) were constructed, in which the intercept represents the arithmetic skill at T1, and the slope represents the rate of linear growth from T1 to T4. The model for arithmetic fluency showed a good fit, $\chi^2 = 4.55, df = 5, p = 0.47, CFI = 1.000, TLI = 1.003, RMSEA = 0.000$, and the correlation between the intercept and slope was not significant (estimated $r = 0.31, p > 0.12$). In turn, the model for arithmetic accuracy did not

fit the data well. The modification indices indicated that the estimated residual of arithmetic accuracy at T3 was related to that of T4, suggesting that the two measurements shared some unique variance that was not included in the model. After incorporating the above relation in the model, the model fit the data very well, $\chi^2 = 7.43, df = 4, p = 0.12, CFI = 0.981, TLI = 0.953, RMSEA = 0.069$, and the correlation between the intercept and slope was significant (estimated $r = 0.82, p < 0.05$). The results also showed a significant variance in the intercepts and slopes of both mathematics skills (for arithmetic accuracy, $\sigma_i^2 = 1.74, p < 0.05, \sigma_s^2 = 1.36, p < 0.05$; for arithmetic fluency, $\sigma_i^2 = 6.29, p < 0.001, \sigma_s^2 = 0.66, p < 0.01$).

Next, six variables at T1 were used to predict the intercept and slope of a linear growth model for each mathematics skill. In both models, the intercept was allowed to correlate with the slope, and the residuals of the predictors were allowed to be correlated. The models predicting growth in arithmetic accuracy and arithmetic fluency are shown in Figures 1, 2, respectively, with non-significant paths removed. Both models fit the data well (for arithmetic accuracy, $\chi^2 = 14.97, df = 16, p = 0.53, CFI = 0.935, TLI = 1.010, RMSEA = 0.000$; for arithmetic fluency, $\chi^2 = 16.63, df = 17, p = 0.48, CFI = 1.000, TLI = 1.002, RMSEA = 0.000$). Nonverbal IQ and speed of processing predicted the intercept of arithmetic accuracy and accounted for 36.4% of the variance. Nonverbal IQ and working memory predicted the slope of arithmetic accuracy and accounted for 31.3% of the variance. Speed of processing, number sense, inhibition and shifting predicted the intercept of arithmetic fluency and accounted for 39.6% of the variance. No variables predicted significantly the slope of arithmetic fluency.

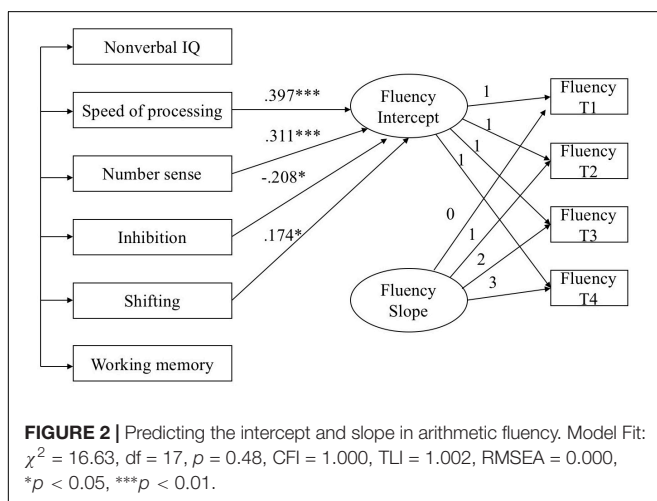
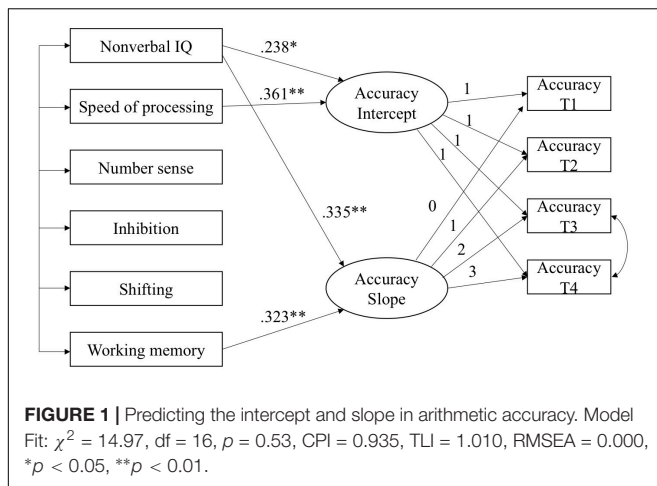
DISCUSSION

This study aimed to examine how the three core EF subcomponents (i.e., inhibition, shifting, and working memory) predict the growth parameters of two mathematics skills (i.e.,

TABLE 2 | Correlations between cognitive predictors and mathematics outcomes.

	1	2	3	4	5	6	7	8	9	10	11	12	13
(1) Nonverbal IQ													
(2) Speed of processing	0.22												
(3) Number sense	0.26	0.39											
(4) Inhibition	-0.25	-0.43	-0.32										
(5) Shifting	-0.21	-0.38	-0.42	0.30									
(6) Working memory	0.23	0.26	0.16	-0.29	-0.07								
(7) Arithmetic accuracy T1	0.23	0.28	0.21	-0.19	-0.09	0.20							
(8) Arithmetic accuracy T2	0.33	0.28	0.26	-0.15	-0.14	0.25	0.44						
(9) Arithmetic accuracy T3	0.36	0.22	0.25	-0.20	-0.17	0.22	0.43	0.55					
(10) Arithmetic accuracy T4	0.33	0.20	0.28	-0.17	-0.21	0.34	0.35	0.51	0.61				
(11) Arithmetic fluency T1	0.06	0.42	0.37	-0.33	-0.14	0.15	0.30	0.38	0.23	0.31			
(12) Arithmetic fluency T2	0.17	0.40	0.30	-0.33	-0.07	0.22	0.34	0.34	0.35	0.29	0.57		
(13) Arithmetic fluency T3	0.19	0.43	0.35	-0.34	-0.16	0.14	0.44	0.31	0.32	0.31	0.60	0.62	
(14) Arithmetic fluency T4	0.18	0.38	0.34	-0.37	-0.13	0.31	0.36	0.36	0.32	0.30	0.54	0.60	0.67

Correlations lower than 0.16 were not significant. Correlations between 0.16 and 0.20 were significant at the 0.05 level and correlations higher than 0.20 were significant at the 0.01 level.



arithmetic accuracy and fluency) in a group of Chinese children followed from Grade 2 to 5. The results showed that the three EF subcomponents were interrelated, but predicted different growth parameters in different mathematics skills. Whereas working memory uniquely predicted the slope of arithmetic accuracy, inhibition and shifting predicted the intercept of arithmetic fluency.

In contrast to our expectation (see Hypothesis 1) and to the findings of some previous studies (e.g., Viterbori et al., 2015; Cragg et al., 2017), working memory did not uniquely predict the growth parameters of arithmetic fluency. This may be due to the fact that neither Cragg et al. (2017) nor Viterbori et al. (2015) controlled for number sense and/or speed of processing before examining the contribution of working memory to arithmetic fluency. However, it may also reflect differences in the amount of working memory involved in the strategies used to solve simple calculations in different countries. Children in North America learn how to solve simple calculations in Grade 1 and by Grade 2 (when we first assessed them) they still use “immature” strategies (e.g., counting on) that tax working memory (e.g., Geary et al., 1996; Bailey et al., 2012; see also Miller et al., 2005, for a review of differences in how children learn mathematics in China and

the United States). In contrast, in China, Grade 2 children solve simple calculations by retrieving the answer from their long-term memory. This is because they have been practicing simple calculations since the age of 3 (when they go to kindergarten).

Second, our results showed that working memory uniquely predicted the slope in arithmetic accuracy (see Geary, 2011; Viterbori et al., 2015, for a similar finding). This suggests that working memory contributes to the learning of new operations, which are basic operations in lower grades and more complex operations in higher grades. Geary (2011) and Yen et al. (2017) also argued that EF may be more important in higher grades, because more complex and difficult operations need the extensive engagement of central executive. Once the operation and calculation reaches an automatic level, working memory may no longer have a role to play in the calculation process (Träff, 2013; Cowan and Powell, 2014).

In line with our second hypothesis, we also found that inhibition uniquely predicted the intercept in arithmetic fluency even after controlling for the effects of nonverbal IQ, speed of processing, and number sense. Viterbori et al. (2015) have argued that inhibition may be involved in the process of retrieving the arithmetic facts and is required for suppressing competing responses. For example, when retrieving the answer 5 in response to $3 + 2$, children need to suppress 6 as the solution to 3×2 , considering that single digit multiplication is learned by most of Chinese children through rote memory (Zhou et al., 2006).

In contrast to our third hypothesis as well as to the findings of previous studies (e.g., Cantin et al., 2016; Cragg et al., 2017; Simanowski and Krajewski, 2017), shifting was a significant predictor of the intercept in arithmetic fluency. A possible explanation may be that we used time scores of shifting, while Cantin et al. (2016) and Cragg et al. (2017) used accuracy scores. It may also be due to the task we used to operationalize arithmetic fluency. Specifically, because BAT mixes up the addition and subtraction problems, children likely had to switch between addition and subtraction mindsets.

However, neither inhibition nor shifting predicted the slope of arithmetic fluency. Because Chinese children learn different calculations when they go to kindergarten (at the age of 3), by the time they reach elementary school they have already mastered simple calculations. Subsequently, when asked to perform simple calculations they rely more on fact retrieval than on procedural strategies (e.g., Geary et al., 1996; Bailey et al., 2012; Vanbinst et al., 2015). Inhibition and shifting may be important in arithmetic fluency in China but in earlier grades when Chinese children learn to perform simple calculations (i.e., the 3 years of kindergarten).

Some limitations of the present study are worth mentioning. First, we used single measures of each EF subcomponent and this may have weakened each construct and subsequently its contribution to mathematics. Future studies should assess each EF component with more tasks. Second, in order to directly compare the contribution of EF subcomponents to timed and untimed mathematics skills, we did not include problem solving since problem solving is a higher-level mathematics skill predicted not only by domain-general skills, but also by reading-related skills (e.g., Andersson, 2008; Träff, 2013).

Third, our measures of working memory and shifting involved numerical stimuli. This may have increased the contribution of their respective constructs to mathematics. However, notice that we controlled for the effects of other cognitive skills that also contained numerical stimuli (e.g., speed of processing, number sense). Fourth, we did not obtain information on family's income. Some studies (e.g., Hackman et al., 2015; Chung et al., 2017) have shown that family's income correlates with both EF and children's math achievement. This implies that the relationship between EF and mathematics might be due to family's income. Future studies should explore this possibility. Fifth, due to time restrictions, we took a purely cognitive view of mathematics. We acknowledge that affective, social, and emotional attributes may play an equally strong role in mathematics development. Finally, although many Chinese parents pay private tutors (typically from commercial education companies) to instruct their children to practice mathematical skills with more homework, we were not able to obtain information on this issue and, as a result, we were not able to control for its effects on mathematics skills.

CONCLUSION

Our study adds to a growing body of research on the contribution of different EF subcomponents to mathematics development (e.g., van der Ven et al., 2012; Van de Weijer-Bergsma et al., 2015; Lee and Bull, 2016) suggesting that different EF subcomponents may contribute to different growth parameters in arithmetic accuracy and fluency, even after controlling for the effects of other known predictors of mathematics (i.e., nonverbal IQ, speed of processing, and number sense). We echo here Cragg and Gilmore's (2014) conclusion that different EF skills contribute to different components of mathematical knowledge as well as Miyake et al.'s (2000) conclusion that the unity of the EF subcomponents is important but it is diversity in what skills

they predict that makes things interesting. From a practical point of view, this suggests that depending on what mathematics outcome we want to predict we should include different types of EF tasks to maximize our predictive power. At the same time, this finding implies that depending on the type of mathematics difficulties a child has (e.g., procedural vs. semantic memory difficulties; Geary, 1993) and to the extent we want to provide an EF intervention (see Dias and Seabra, 2017), we need to focus on different EF subcomponents to maximize our chances to be effective.

AUTHOR CONTRIBUTIONS

CD, GG, WW, and AT designed the study. WW and LG collected the data, prepared the data for analysis, and wrote the manuscript. All authors interpreted the data and discussed the results. GG, CD, WW, and LG revised the manuscript.

FUNDING

This study was supported by a grant from Shanghai Normal University (A-7031-18-004025), a grant from the National Natural Science Foundation of China (Grant No. 71373081), and a grant from China Postdoctoral Science Foundation (2016M601624).

ACKNOWLEDGMENTS

We would like to thank Xue Shell Jing, Liu Hai-Lun, Wang Lei, and Gao Shu-Xian at East China Normal University for their assistance in collecting the data for this study, and Dr. Tomohiro Inoue at University of Alberta for his assistance with the data analysis.

REFERENCES

- Agostino, A., Johnson, J., and Pascual-Leone, J. (2010). Executive functions underlying multiplicative reasoning: problem type matters. *J. Exp. Child Psychol.* 105, 286–305. doi: 10.1016/j.jecp.2009.09.006
- Allan, N. P., Hume, L. E., Allan, D. M., Farrington, A. L., and Lonigan, C. J. (2014). Relations between inhibitory control and the development of academic skills in preschool and kindergarten: a meta-analysis. *Dev. Psychol.* 50, 2368–2379. doi: 10.1037/a0037493
- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychol.* 8, 71–82. doi: 10.1076/chin.8.2.71.8724
- Andersson, U. (2008). Working memory as a predictor of written arithmetical skills in children: the importance of central executive functions. *Br. J. Educ. Psychol.* 78, 181–203. doi: 10.1348/000709907X209854
- Bailey, D. H., Littlefield, A., and 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. *J. Exp. Child Psychol.* 113, 78–92. doi: 10.1016/j.jecp.2012.04.014
- Best, J. R., Miller, P. H., and Jones, L. L. (2009). Executive functions after age 5: changes and correlates. *Dev. Rev.* 29, 180–200. doi: 10.1016/j.dr.2009.05.002
- Blair, C., and Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Dev.* 78, 647–663. doi: 10.1111/j.1467-8624.2007.01019.x
- Bull, R., Espy, K. A., and Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: longitudinal predictors of mathematical achievement at age 7 years. *Dev. Neuropsychol.* 33, 205–228. doi: 10.1080/87565640801982312
- Bull, R., and Lee, K. (2014). Executive functioning and mathematics achievement. *Child Dev. Perspect.* 8, 36–41. doi: 10.1111/cdep.12059
- Bull, R., and Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: inhibition, switching, and working memory. *Dev. Neuropsychol.* 19, 273–293. doi: 10.1207/S15326942DN1903_3
- Campbell, J. I., and Xue, Q. (2001). Cognitive arithmetic across cultures. *J. Exp. Psychol. Gen.* 130, 299–315. doi: 10.1037/0096-3445.130.2.299
- Cantin, R. H., Gnaedinger, E. K., Gallaway, K. C., Hesson-McInnis, M. S., and Hund, A. M. (2016). Executive functioning predicts reading, mathematics, and theory of mind during the elementary years. *J. Exp. Child Psychol.* 146, 66–78. doi: 10.1016/j.jecp.2016.01.014
- Chan, R. C., Shum, D., Touloupoulou, T., and Chen, E. Y. (2008). Assessment of executive functions: review of instruments and identification of critical issues. *Arch. Clin. Neuropsychol.* 23, 201–216. doi: 10.1016/j.acn.2007.08.010
- Chung, K. K. H., Liu, H., McBride, C., Wong, A. M. Y., and Lo, J. C. (2017). How socioeconomic status, executive functioning and verbal interactions contribute

- to early academic achievement in Chinese children. *Educ. Psychol.* 37, 402–420. doi: 10.1080/01443410.2016.1179264
- Clark, C. A., Pritchard, V. E., and Woodward, L. J. (2010). Preschool executive functioning abilities predict early mathematics achievement. *Dev. Psychol.* 46, 1176–1191. doi: 10.1037/a0019672
- Cowan, R., and Powell, D. (2014). The contributions of domain-general and numerical factors to third-grade arithmetic skills and mathematical learning disability. *J. Educ. Psychol.* 106, 214–229. doi: 10.1037/a0034097
- Cragg, L., and Gilmore, C. (2014). Skills underlying mathematics: the role of executive function in the development of mathematics proficiency. *Trends Neurosci. Educ.* 3, 63–68. doi: 10.1016/j.tine.2013.12.001
- Cragg, L., Keeble, S., Richardson, S., Roome, H. E., and Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition* 162, 12–26. doi: 10.1016/j.cognition.2017.01.014
- Cui, J., Georgiou, G. K., Zhang, Y., Li, Y., Shu, H., and Zhou, X. (2017). Examining the relationship between rapid automatized naming and arithmetic fluency in Chinese kindergarten children. *J. Exp. Child Psychol.* 154, 146–163. doi: 10.1016/j.jecp.2016.10.008
- Day, S. L., Connor, C. M., and McClelland, M. M. (2015). Children's behavioral regulation and literacy: the impact of the first grade classroom environment. *J. Sch. Psychol.* 53, 409–428. doi: 10.1016/j.jsp.2015.07.004
- Deng, C. P., Liu, M., Wei, W., Chan, R. C., and Das, J. P. (2011). Latent factor structure of the Das-Naglieri cognitive assessment system: a confirmatory factor analysis in a Chinese setting. *Res. Dev. Disabil.* 32, 1988–1997. doi: 10.1016/j.ridd.2011.04.005
- Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750
- Dias, N. M., and Seabra, A. G. (2017). Intervention for executive functions development in early elementary school children: effects on learning and behaviour, and follow-up maintenance. *Educ. Psychol.* 37, 468–486. doi: 10.1080/01443410.2016.1214686
- Espy, K. A., McDiarmid, M. M., Cwik, M. F., Stalets, M. M., Hamby, A., and Senn, T. E. (2004). The contribution of executive functions to emergent mathematical skills in preschool children. *Dev. Neuropsychol.* 26, 465–486. doi: 10.1207/s15326942dn2601_6
- Friso-van den Bos, I., van der Ven, S. H. G., Kroesbergen, E. H., and van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: a meta-analysis. *Educ. Res. Rev.* 10, 29–44. doi: 10.1016/j.edurev.2013.05.003
- Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *J. Educ. Psychol.* 98, 29–43. doi: 10.1037/0022-0663.98.1.29
- Fuhs, M. W., Hornburg, C. B., and McNeil, N. M. (2016). Specific early number skills mediate the association between executive functioning skills and mathematics achievement. *Dev. Psychol.* 52, 1217–1235. doi: 10.1037/dev0000145
- Geary, D. C. (1993). Mathematical disabilities cognitive, neuropsychological, and genetic components. *Psychol. Bull.* 114, 345–362. doi: 10.1037/0033-2909.114.2.345
- Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Dev. Psychol.* 47, 1539–1552. doi: 10.1037/a0025510
- Geary, D. C., Bailey, D. H., Littlefield, A., Wood, P., Hoard, M. K., and Nugent, L. (2009). First-grade predictors of mathematical learning disability: a latent class trajectory analysis. *Cogn. Dev.* 24, 411–429. doi: 10.1016/j.cogdev.2009.10.001
- Geary, D. C., Bow-Thomas, C. C., Liu, F., and Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: influence of age, language, and schooling. *Child Dev.* 67, 2022–2044. doi: 10.2307/1131607
- Georgiou, G., and Das, J. P. (2018). Direct and indirect effects of executive function on reading comprehension in young adults. *J. Res. Read.* 41, 243–258. doi: 10.1111/1467-9817.12091
- Gersten, R., and Chard, D. (1999). Number sense: rethinking arithmetic instruction for students with mathematical disabilities. *J. Spec. Educ.* 33, 18–28. doi: 10.1177/002246699903300102
- Gilmore, C., Keeble, S., Richardson, S., and Cragg, L. (2015). The role of cognitive inhibition in different components of arithmetic. *ZDM* 47, 771–782. doi: 10.1007/s11858-014-0659-y
- Hackman, D. A., Gallop, R., Evans, G. W., and Farah, M. J. (2015). Socioeconomic status and executive function: developmental trajectories and mediation. *Dev. Sci.* 18, 686–702. doi: 10.1111/desc.12246
- Imbo, I., and LeFevre, J. A. (2009). Cultural differences in complex addition: efficient Chinese versus adaptive Belgians and Canadians. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1465–1476. doi: 10.1037/a0017022
- Kolkman, M. E., Hoijtink, H. J., Kroesbergen, E. H., and Leseman, P. P. (2013). The role of executive functions in numerical magnitude skills. *Learn. Individ. Differ.* 24, 145–151. doi: 10.1016/j.lindif.2013.01.004
- Lan, X., Legare, C. H., Ponitz, C. C., Li, S., and Morrison, F. J. (2011). Investigating the links between the subcomponents of executive function and academic achievement: a cross-cultural analysis of Chinese and American preschoolers. *J. Exp. Child Psychol.* 108, 677–692. doi: 10.1016/j.jecp.2010.11.001
- Lee, K., and Bull, R. (2016). Developmental changes in working memory, updating, and math achievement. *J. Educ. Psychol.* 108, 869–882. doi: 10.1037/edu0000090
- Lee, K., Ng, S. F., Pe, M. L., Ang, S. Y., Hasshim, M. N., and Bull, R. (2012). The cognitive underpinnings of emerging mathematical skills: executive functioning, patterns, numeracy, and arithmetic. *Br. J. Educ. Psychol.* 82, 82–99. doi: 10.1111/j.2044-8279.2010.02016.x
- Lehto, J. E., Juujärvi, P., Kooistra, L., and Pulkkinen, L. (2003). Dimensions of executive functioning: evidence from children. *Br. J. Educ. Psychol.* 21, 59–80. doi: 10.1348/02615100321164627
- Liao, C. H., Georgiou, G. K., and Parrila, R. (2008). Rapid naming speed and Chinese character recognition. *Read. Writ.* 21, 231–253. doi: 10.1007/s11145-007-9071-0
- Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., and Lindberg, S. (2016). Differences in arithmetic performance between Chinese and German children are accompanied by differences in processing of symbolic numerical magnitude. *Front. Psychol.* 7:1337. doi: 10.3389/fpsyg.2016.01337
- Lubin, A., Regrin, E., Boulch, L., Pacton, S., and Lanoë, C. (2016). Executive functions differentially contribute to fourth graders' mathematics, reading, and spelling skills. *J. Cogn. Educ. Psychol.* 15, 444–463. doi: 10.1891/1945-8959.15.3.444
- McClelland, M. M., Cameron, C. E., Duncan, R., Bowles, R. P., Acoc, A. C., Miao, A., et al. (2014). Predictors of early growth in academic achievement: the head-toes-knees-shoulders task. *Front. Psychol.* 5:599. doi: 10.3389/fpsyg.2014.00599
- Miller, K. F., Kelly, M., and Zhou, X. (2005). "Learning mathematics in China and the United States: cross-cultural insights into the nature and course of mathematical development," in *Handbook of Mathematical Cognition*, ed. J. I. D. Campbell (New York, NY: Psychology Press), 163–178.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., and Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 49, 49–100. doi: 10.1006/cogp.1999.0734
- Monette, S., Bigras, M., and Guay, M. C. (2011). The role of the executive functions in school achievement at the end of Grade 1. *J. Exp. Child Psychol.* 109, 158–173. doi: 10.1016/j.jecp.2011.01.008
- Naglieri, J. A., and Das, J. P. (1997). *Das-Naglieri Cognitive Assessment System*. Itasca, IL: Riverside Publishing.
- Peng, P., Namkung, J., Barnes, M., and Sun, C. (2015). A meta-analysis of mathematics and working memory: moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *J. Educ. Psychol.* 108, 455–473. doi: 10.1037/edu0000079
- Peng, P., Sun, C., Li, B., and Tao, S. (2012). Phonological storage and executive function deficits in children with mathematics difficulties. *J. Exp. Child Psychol.* 112, 452–466. doi: 10.1016/j.jecp.2012.04.004
- Purpura, D. J., Schmitt, S. A., and Ganley, C. M. (2017). Foundations of mathematics and literacy: the role of executive functioning components. *J. Exp. Child Psychol.* 153, 15–34. doi: 10.1016/j.jecp.2016.08.010
- Raghubar, K. P., Barnes, M. A., and Hecht, S. A. (2010). Working memory and mathematics: a review of developmental, individual difference, and cognitive approaches. *Learn. Individ. Differ.* 20, 110–122. doi: 10.1016/j.lindif.2009.10.005
- Reitan, R. M., and Wolfson, D. (1992). Conventional intelligence measurements and neuropsychological concepts of adaptive abilities. *J. Clin. Psychol.* 48, 521–529. doi: 10.1002/1097-4679(199207)48:4<521::AID-JCLP2270480414>3.0.CO;2-C

- Rose, S. A., Feldman, J. F., and Jankowski, J. J. (2011). Modeling a cascade of effects: the role of speed and executive functioning in preterm/full-term differences in academic achievement. *Dev. Sci.* 14, 1161–1175. doi: 10.1111/j.1467-7687.2011.01068.x
- Siegler, R. S., and Shrager, J. (1984). "Strategy choices in addition and subtraction: how do children know what to do," in *Origins of Cognitive Skills*, ed C. Sophian (Hillsdale, NJ: Erlbaum), 229–293.
- Simanowski, S., and Krajewski, K. (2017). Specific preschool executive functions predict unique aspects of mathematics development: a 3-year longitudinal study. *Child Dev.* doi: 10.1111/cdev.12909 [Epub ahead of print].
- Stevenson, H. W., Lee, S. Y., Chen, C., Lummis, M., Stigler, J., Fan, L., et al. (1990). Mathematics achievement of children in China and the United States. *Child Dev.* 61, 1053–1066. doi: 10.2307/1130875
- Swanson, H. L. (2006). Cross-sectional and incremental changes in working memory and mathematical problem solving. *J. Educ. Psychol.* 98, 265–281. doi: 10.1037/0022-0663.98.2.265
- Tabachnick, B., and Fidell, L. (2007). *Experimental Designs Using ANOVA*. Belmont, CA: Thomson Brooks/Cole.
- Testa, R., Bennett, P., and Ponsford, J. (2012). Factor analysis of nineteen executive function tests in a healthy adult population. *Arch. Clin. Neuropsychol.* 27, 213–224. doi: 10.1093/arcclin/acr112
- Thorell, L. B., Veleiro, A., Siu, A. F., and Mohammadi, H. (2013). Examining the relation between ratings of executive functioning and academic achievement: findings from a cross-cultural study. *Child Neuropsychol.* 19, 630–638. doi: 10.1080/09297049.2012.727792
- Träff, U. (2013). The contribution of general cognitive abilities and number abilities to different aspects of mathematics in children. *J. Exp. Child Psychol.* 116, 139–156. doi: 10.1016/j.jecp.2013.04.007
- Van de Weijer-Bergsma, E., Kroesbergen, E. H., and Van Luit, J. E. (2015). Verbal and visual-spatial working memory and mathematical ability in different domains throughout primary school. *Mem. Cognit.* 43, 367–378. doi: 10.3758/s13421-014-0480-4
- van der Sluis, S., de Jong, P. F., and van der Leij, A. (2007). Executive functioning in children, and its relations with reasoning, reading, and arithmetic. *Intelligence* 35, 427–449. doi: 10.1016/j.intell.2006.09.001
- van der Ven, S. H., Kroesbergen, E. H., Boom, J., and Leseman, P. P. (2012). The development of executive functions and early mathematics: a dynamic relationship. *Br. J. Educ. Psychol.* 82, 100–119. doi: 10.1111/j.2044-8279.2011.02035.x
- Vanbinst, K., Ceulemans, E., Ghesquière, P., and De Smedt, B. (2015). Profiles of children's arithmetic fact development: a model-based clustering approach. *J. Exp. Child Psychol.* 133, 29–46. doi: 10.1016/j.jecp.2015.01.003
- Vandenbroucke, L., Verschueren, K., and Baeyens, D. (2017). The development of executive functioning across the transition to first grade and its predictive value for academic achievement. *Learn. Instr.* 49, 103–112. doi: 10.1016/j.learninstruc.2016.12.008
- Viterbori, P., Usai, M. C., Traverso, L., and De Franchis, V. (2015). How preschool executive functioning predicts several aspects of math achievement in Grades 1 and 3: a longitudinal study. *J. Exp. Child Psychol.* 140, 38–55. doi: 10.1016/j.jecp.2015.06.014
- Wang, J., and Lin, E. (2009). A meta-analysis of comparative studies on Chinese and US students' mathematics performance: implications for mathematics education reform and research. *Educ. Res. Rev.* 4, 177–195. doi: 10.1016/j.edurev.2009.06.003
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. New York, NY: Psychological Corporation.
- Wechsler, D. L. (2002). *Wechsler Individual Achievement Test-Second Edition (WIAT-II)*. San Antonio, TX: The Psychological Corporation.
- Willoughby, M. T., Blair, C. B., Wirth, R. J., Greenberg, M., and The Family Life Project Investigators. (2012). The measurement of executive function at age 5: psychometric properties and relationship to academic achievement. *Psychol. Assess.* 24, 226–239. doi: 10.1037/a0025361
- Woodcock, R. W., and Johnson, M. B. (1989). *Tests of Cognitive Ability*. New York, NY: Riverside Publishing.
- Wu, K. K., Chan, S. K., Leung, P. W., Liu, W. S., Leung, F. L., and Ng, R. (2011). Components and developmental differences of executive functioning for school-aged children. *Dev. Neuropsychol.* 36, 319–337. doi: 10.1080/87565641.2010.549979
- Xu, F., Han, Y., Sabbagh, M. A., Wang, T., Ren, X., and Li, C. (2013). Developmental differences in the structure of executive function in middle childhood and adolescence. *PLoS One* 8:e77770. doi: 10.1371/journal.pone.0077770
- Yen, M. H., Han, C. C., Yu, P. C., Yang, T. H., Didino, D., Butterworth, B., et al. (2017). The influence of memory updating and number sense on junior high school math attainment. *Learn. Individ. Differ.* 54, 30–40. doi: 10.1016/j.lindif.2017.01.012
- Yeniad, N., Malda, M., Mesman, J., van IJzendoorn, M. H., and Pieper, S. (2013). Shifting ability predicts math and reading performance in children: a meta-analytical study. *Learn. Individ. Differ.* 23, 1–9. doi: 10.1016/j.lindif.2012.10.004
- Zhang, X. (2016). Linking language, visual-spatial, and executive function skills to number competence in very young Chinese children. *Early Child. Res. Q.* 36, 178–189. doi: 10.1016/j.ecresq.2015.12.010
- Zhang, X., Hu, B. Y., Ren, L., and Fan, X. (2017). Pathways to reading, mathematics, and science: examining domain-general correlates in young Chinese children. *Contemp. Educ. Psychol.* 51, 366–377. doi: 10.1016/j.cedpsych.2017.09.004
- Zhou, X., Chen, C., Dong, Q., Zhang, H., Zhou, R., Zhao, H., et al. (2006). Event-related potentials of single-digit addition, subtraction, and multiplication. *Neuropsychologia* 44, 2500–2507. doi: 10.1016/j.neuropsychologia.2006.04.003

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

Copyright © 2018 Wei, Guo, Georgiou, Tavouktsoglou and Deng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Children's Non-symbolic and Symbolic Numerical Representations and Their Associations With Mathematical Ability

Yanjun Li^{1,2}, Meng Zhang³, Yinghe Chen^{1*}, Zhijun Deng¹, Xiaoshuang Zhu¹ and Shijia Yan⁴

¹ School of Developmental Psychology, Faculty of Psychology, Beijing Normal University, Beijing, China, ² National Innovation Center for Assessment of Basic Education Quality, Beijing Normal University, Beijing, China, ³ Department of Psychology, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States, ⁴ China Aerospace Academy of Systems Science and Engineering, Institute of Information Control, China Aerospace Science and Technology Corporation, Beijing, China

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

Pom Charras,
Centre National de la Recherche
Scientifique (CNRS), France
Robert Reeve,
University of Melbourne, Australia

*Correspondence:

Yinghe Chen
chenyinghe@bnu.edu.cn

Specialty section:

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

Received: 05 March 2018

Accepted: 01 June 2018

Published: 25 June 2018

Citation:

Li Y, Zhang M, Chen Y, Deng Z, Zhu X
and Yan S (2018) Children's
Non-symbolic and Symbolic
Numerical Representations and Their
Associations With Mathematical
Ability. *Front. Psychol.* 9:1035.
doi: 10.3389/fpsyg.2018.01035

Most empirical evidence supports the view that non-symbolic and symbolic representations are foundations for advanced mathematical ability. However, the detailed development trajectories of these two types of representations in childhood are not very clear, nor are the different effects of non-symbolic and symbolic representations on the development of mathematical ability. We assessed 253 4- to 8-year-old children's non-symbolic and symbolic numerical representations, mapping skills, and mathematical ability, aiming to investigate the developmental trajectories and associations between these skills. Our results showed non-symbolic numerical representation emerged earlier than the symbolic one. Four-year-olds were capable of non-symbolic comparisons but not symbolic comparisons; five-year-olds performed better at non-symbolic comparisons than symbolic comparisons. This performance difference disappeared at age 6. Children at age 6 or older were able to map between symbolic and non-symbolic quantities. However, as children learn more about the symbolic representation system, their advantage in non-symbolic representation disappeared. Path analyses revealed that a direct effect of children's symbolic numerical skills on their math performance, and an indirect effect of non-symbolic numerical skills on math performance via symbolic skills. These results suggest that symbolic numerical skills are a predominant factor affecting math performance in early childhood. However, the influences of symbolic and non-symbolic numerical skills on mathematical performance both declines with age.

Keywords: non-symbolic numerical representation, symbolic numerical representation, mapping, mathematical ability, mathematical development

INTRODUCTION

The Developmental Trajectories of Non-symbolic and Symbolic Representation Abilities

A variety of studies have suggested that animals and humans shared the capacity of non-symbolic representation (Wynn, 1992; Pica et al., 2004; Flombaum et al., 2005), which has been attributed to the so-called approximate number system (ANS) (Feigenson et al., 2004; Barth et al., 2005, 2006, 2008; Dehaene, 2011). The ANS system has three features. First, it is inherent and universal

(Wynn, 1992; Pica et al., 2004; Flombaum et al., 2005); animals and humans share the system. Second, it represents quantities in an approximate way (Feigenson et al., 2004). Third, the precision of ANS system increases with age (Halberda et al., 2008). Correspondingly, there are three different characteristics for symbolic number representation system. First, it is an acquired system, it is affected by the language faculties (Pica et al., 2004; Xenidou-Dervou et al., 2015). Second, it represents quantities precisely (Izard and Dehaene, 2008; Mussolin et al., 2014). Third, with age, the system can manipulate increasingly larger range with higher accuracy (Halberda et al., 2008; Praet and Desoete, 2014).

Children's non-symbolic skills emerge early and develop continuously over time (Barth et al., 2005, 2006, 2008; Halberda et al., 2008). Libertus et al. (2011) assessed non-symbolic skills with numbers range 4 to 15. They found that 4-year-olds were able to complete their non-symbolic comparison task. Toll et al. (2015) tested non-symbolic skills with a larger range of 1–100 and found the similar results in 4-year-olds. Wagner and Johnson (2011) assessed non-symbolic skills with numbers range 1–50. They found 3-year-olds performed above chance level in non-symbolic comparison task with numerosities 1–4. Many studies (Barth et al., 2005; Sasanguie et al., 2013; Hyde et al., 2014; Vanbinst et al., 2015) examined the non-symbolic comparison ability in 5-year-olds and older children. They found the skill kept developing during childhood, even till adulthood. Barth et al. (2005) found that adults were significantly more accurate than 5-year-old children in the non-symbolic comparison task.

Research showed symbolic skills emerged at 5 years old, before the start of formal schooling (Kolkman et al., 2013). Children were able to do symbolic representation task at age 5 (Gillmore et al., 2007). What makes them capable of symbolic numerical representations before formally learning numerical symbols? Some researchers (Gillmore et al., 2007) argued that children might pass the task with the help of their ANS. It is plausible that they converted symbolic Arabic numbers to non-symbolic numerosities. In other word, they had the mapping ability, which enabled the process of transforming non-symbolic representation and symbolic representation information into one another. Other researchers argued that informal mathematical activities help improve children's symbolic skills (Skwarchuk et al., 2014; Berkowitz et al., 2015). Although 4- or 5-year-old children have not obtained mathematical education from school, they may have already been exposed to many informal mathematic activities, such as playing number board game, reading stories involved quantities, and so on. With so many possible exposures to mathematical knowledge, this study tries to explore whether children as young as 4 years old are able to represent and compare symbolic Arabic numbers.

The relationship between symbolic skills and non-symbolic skills has been discussed a lot in this field. Some researchers claim that non-symbolic and symbolic skills are separable (They adopted non-symbolic comparison and symbolic comparison tasks which are similar to tasks in our current study) (Kolkman et al., 2013). They rely on two distinct systems and do not share the same underlying ability (Xenidou-Dervou et al., 2015). Other researchers believed that both non-symbolic and

symbolic comparison abilities, to some extent, relied on the ANS system (Chen and Li, 2014; van Marle et al., 2014). Furthermore, the majority of previous studies focused on the correlation between non-symbolic and symbolic representation skills (Castronovo and Goebel, 2012; Gobel et al., 2014). Most researchers believe there is a positive correlation between non-symbolic and symbolic skills (Kolkman et al., 2013; van Marle et al., 2014; Toll et al., 2015). Other researchers (Fazio et al., 2014) found no correlation between these two types of skills. The available evidence is not congruent, both distinctions and connections between symbolic and non-symbolic comparison abilities were reported. The development trajectories of these two are not very clear. Some tasks used by previous researchers were too difficult to detect children's emerging numerical skills. For example, Xenidou-Dervou et al. (2015) assessed 5- and 6-year-olds' non-symbolic and symbolic abilities by using approximate addition tasks, which were harder than comparison. In their task, children had to add the two quantities first and then to compare the numerosities. That is to say, their task also required children's arithmetic ability at the same time. The present study used comparison tasks to test both symbolic and non-symbolic abilities. We aim to provide more comprehensive developmental trajectories of non-symbolic and symbolic capacities in preschoolers and young primary students.

The Associations Between Numerical Representation Skills and Mathematical Ability

The association between non-symbolic representation and mathematical ability is not clear. Many studies showed positive correlations between non-symbolic representation skills and mathematical ability in children and adults (DeWind and Brannon, 2012; Libertus et al., 2012; Bonny and Lourenco, 2013). Libertus et al. (2012) assessed 3- to 5-year-olds' non-symbolic comparison precision and mathematical ability. They found there was a significant positive correlation between the precision of non-symbolic task and mathematical achievement. Halberda et al. (2008) found similar results in older children. Furthermore, longitudinal data showed that non-symbolic skills in early childhood significantly predicted later mathematical abilities (Halberda et al., 2008; Mazzocco et al., 2011; Libertus et al., 2013). However, other researchers did not find positive correlations between non-symbolic representation skills and mathematical ability in children (Holloway and Ansari, 2009; Sasanguie et al., 2013) and adults (Inglis et al., 2011; Price et al., 2012). It appears that not all researchers consider that non-symbolic representation ability and mathematical ability are related. Therefore, the issue, whether the ability of non-symbolic representation play an important role in the development of mathematical ability or not, needs further explorations.

Researchers have reached a consensus about the relationship between symbolic skills and mathematical ability. That is, symbolic skills have a significant impact on mathematical ability. Bugden and Ansari (2011) found a significant positive correlation between symbolic comparison skills and mathematical ability in 1st and 2nd grade children from primary school. Toll et al. (2015)

investigated children's non-symbolic and symbolic comparison skills; they found that symbolic comparison skills were the most important predictor for mathematical ability. Similar results were also found from a longitudinal study (Kolkman et al., 2013).

Most empirical evidence supports the view that non-symbolic and symbolic comparison skills are foundations for advanced mathematical ability (Libertus et al., 2011; Castronovo and Goebel, 2012). van Marle et al. (2014) assessed non-symbolic, symbolic skills, and mathematics ability of 4-year-olds. They found the relation between non-symbolic skills and mathematics ability was completely mediated by children's performance on the symbolic comparison task. Similar results were also found in 6-year-olds (Gobel et al., 2014). However, Fazio et al. (2014) assessed children of 10 years old, and they found symbolic and non-symbolic skills related to mathematics ability uniquely. Up to now, it still seems unclear how non-symbolic, symbolic comparison skills, and mathematical performance relate to each other.

In addition, some researchers believed that the ability to map between symbolic and non-symbolic quantities was an important factor in the development of children's mathematical ability (Brankaer et al., 2014). This may be because the mapping capability reflects an individual's ability to process different types of magnitude information. The better one is at mapping, the better he/she could learn advanced mathematics. Mundy and Gilmore (2009) tested children's bi-directional mapping ability and their mathematical performance. A significant prediction of mapping ability was found for mathematical performance. Similar results were also found by Kolkman et al.'s (2013) and Brankaer et al. (2014) path analyses. However, Friso-van Den Bos et al. (2015) tracked 442 5-year-olds for 3 years; they found children's mapping skill did not significantly predict their mathematical achievements. Therefore, the impact of mapping skills on mathematical ability has not been uniformly concluded.

Present Study

In sum, this study aims to achieve two goals. First, we aim to provide detailed development trajectories of non-symbolic and symbolic representation skills in childhood. Previous studies mostly focused a few age groups (Barth et al., 2005, 2006, 2008; Gilmore et al., 2007; Xenidou-Dervou et al.'s 2015). Data capturing a longer developmental period throughout childhood are needed. The available evidence showed both distinctions and connections between symbolic and non-symbolic comparison abilities. We predict that children are more experienced at the non-symbolic task than symbolic task in early childhood, but as they learn more about the symbolic representation system, children's advantage in non-symbolic skill will disappear. Second, this study aims to investigate the associations between numerical representation skills and mathematical ability in childhood. Researchers investigating the issue focused on different age ranges and therefore generated different results (Gobel et al., 2014; van Marle et al., 2014; Friso-van Den Bos et al., 2015). The exact relations between non-symbolic, symbolic comparison, and mathematical performance remain unclear. We focused the age range of 4 to 8 and predicted that the relationships between these

three types of abilities might be different for different age groups in our study.

MATERIALS AND METHODS

Ethics Statement

This research was approved by the local ethical committee of Beijing Normal University. We obtained informed written consent from caretakers or guardians on behalf of the child participants involved in the study, according to the institutional guidelines of Beijing Normal University.

Participants

A total of 253 children (116 girls) were recruited from 2 public schools located in Baoji, Shaanxi province, China. Forty-six 4-year-olds ($M = 48.1$ months, $SD = 4.2$), 61 5-year-olds ($M = 59.6$ months, $SD = 3.6$), and 62 6-year-olds ($M = 73.4$ months, $SD = 3.4$) were recruited from one kindergarten; 39 7-year-olds ($M = 83.2$ months, $SD = 2.5$) and 45 8-year-olds ($M = 96.3$ months, $SD = 3.4$) were recruited from a primary school (the 1st and 2nd grades). All children were tested around March, during the second half of the Chinese academic year. All children are Mandarin native speakers. They were mostly from families of middle socioeconomic status. All children gave oral consent and their parents gave written consent before participation. A gift (i.e., a book) was sent to each child after participation.

Measures

Number-Naming

Children's number-naming ability was measured. They were asked to read loudly 50 Arabic numbers, which were written in five lines on a piece of paper (21 cm $\times$ 29.7 cm). Numbers on the five lines were 1–10, 11–20, 21–30, 31–40, and 41–50 successively. Children obtained 1 point for successfully naming all numbers in one line. Otherwise, they obtained 0 point. The total scores ranged from 0 to 5.

Verbal-Counting

To assess verbal-counting skills, children were asked to count loudly numbers from 1 to 100. These numbers were divided to ten groups (i.e., 1–9, 10–19, 20–29, until 100). They obtained 1 point for successfully counting one entire group. Otherwise, they obtained 0 point. The total scores ranged from 0 to 10.

Non-symbolic Comparison

We tested children's non-symbolic skills using tasks programmed in E-prime. Similar to Wagner and Johnson (2011), we presented participants two black dots arrays and they were asked to decide, without counting, which one contained more dots (see **Figure 1A**). Children were instructed pressing "C" key for quantity on the left and pressing "M" key for quantity on the right. They had a maximum of 10 s to respond and they were required to respond as accurately and quickly as possible. If children did not respond within the 10 s, the trial would automatically be coded as incorrect. The inter-trial interval was

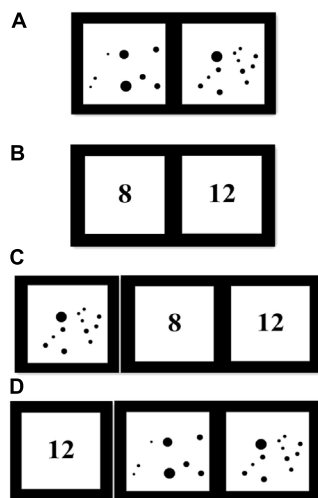


FIGURE 1 | Schematic depictions of numerical comparison tasks and mapping tasks. **(A)** An example trial of non-symbolic comparison task. **(B)** An example trial of symbolic comparison task. **(C)** An example trial of non-symbolic to symbolic mapping task. **(D)** An example trial of symbolic to non-symbolic mapping task.

1000 ms. All children received four practice trials, followed by feedback (“√” or “×”) to make sure they understand the task. After that, they received 32 test trials without feedback.

The numerosities included in this task ranged from 5 to 50. The numerical ratios between the two dot arrays were 2/3, 3/4, 4/5, 5/6. There were eight test trials at each ratio level¹. The order of test trials was random. The probability of large or small numerosities is balanced on the two sides. The dots were constructed in Microsoft Visual C++ 6.0, with the size ranging from 0.2 to 0.6 cm. To rule out judgments based on the continuous dimension of surface area rather than number, the paired dot arrays were matched for total area filled (Feigenson et al., 2002; Rousselle et al., 2004).

Symbolic Comparison

This task was identical to the non-symbolic comparison task except that all dots were replaced by their corresponding Arabic numbers (see **Figure 1B**). Numbers used in each comparison were the same as those in the non-symbolic task. All children received 4 practice trials and 32 test trials.

Mapping

We used a similar task to Mundy and Gilmore (2009), which contained two sub-tasks: (1) Non-symbolic to symbolic mapping task (N-S task). In this task, a target dot array was presented, followed by two alternative Arabic numbers (See **Figure 1C**). Children were asked, “Which Arabic number was equal with the

previous dot array?” (2) Symbolic to non-symbolic mapping task (S-N task). In this task, a target Arabic number was presented, followed by two alternative dot arrays (See **Figure 1D**). Children were asked, “Which dot array was equal with the previous Arabic number?” similarly, children were asked to press “C” or “M” key to response. The target quantity lasted for 1000 ms and then the alternative choices were presented. Children had a maximum of 10 s to respond and they were required to respond as accurately and quickly as possible. If children did not respond within the 10 s, the trial would automatically be coded as incorrect. The inter-trial interval was 1000 ms. For sub-tasks, children received 4 practice trials and 24 test trials.

The target quantities varied from 5 to 50, and the alternative choices consisted of the correct quantity and a distractor. The ratio between the correct quantity and the distractor were 2/3 and 4/5. There were 12 test trials at each ratio level². The correct quantities were counterbalanced in comparable amount within a pair (i.e., larger or smaller) across trials. The same numerosities were tested in both sub-tasks.

Mathematical Competence

We administered Form A of the Test of Early Mathematics Ability-Third Edition (TEMA-3; Ginsburg and Baroody, 2003) to assess their mathematical ability. The TEMA-3 measures many aspects of mathematical performance in childhood, such as numeracy skills (e.g., verbally naming written numbers), number-comparison skills (e.g., determining which of two dot arrays is more), calculation skills (e.g., solving addition or subtraction problems physically or mentally), and number concepts (e.g., answering how many hundreds are in one thousand). It consists of 72 items. Following the standardized administration of the TEMA-3, we started testing with items according the norms for each age group. The test stopped when a child answered 5 consecutive items incorrectly. Scores from the TEMA-3 was normalized for children from 3 years 0 months to 8 years 11 months, and previous research (Ginsburg and Baroody, 2003; Mazzocco et al., 2011) showed relatively high test-retest reliabilities ($r = 0.82, 0.93$) of TEMA-3. Meanwhile, children's performances on TEMA-3 are also highly correlated with their performances on other math achievement tests (Newcomer, 2001; Woodcock et al., 2001).

Procedure

Children were tested individually in a quiet laboratory room, accompanied by one experimenter. All participants complete the number-naming and verbal-counting tasks first, and then the non-symbolic, symbolic comparison tasks and mapping task, which were programmed in E-prime version 2.0 (Psychological Software Tools, Pittsburgh, PA, United States) and presented by a Dell E450 computer. Children complete TEMA-3 last. A short break was provided in-between of tasks. Children received a small reward after the experiment.

¹The paired arrays tested for ratio 2/3 were 6 vs. 9, 8 vs. 12, 10 vs. 15, 12 vs. 18, 14 vs. 21, 16 vs. 24, 18 vs. 27, and 20 vs. 30. The paired arrays tested for ratio 3/4 were 6 vs. 8, 9 vs. 12, 12 vs. 16, 15 vs. 20, 18 vs. 24, 21 vs. 28, 24 vs. 32, and 27 vs. 36. The paired arrays tested for ratio 4/5 were 8 vs. 10, 12 vs. 15, 16 vs. 20, 20 vs. 25, 24 vs. 30, 28 vs. 35, 32 vs. 40, and 36 vs. 45. The paired arrays tested for ratio 5/6 were 5 vs. 6, 10 vs. 12, 15 vs. 18, 20 vs. 24, 25 vs. 30, 30 vs. 36, 35 vs. 42, and 40 vs. 48.

²The pairs tested for ratio 2/3 were 6 vs. 9, 8 vs. 12, 10 vs. 15, 12 vs. 18, 14 vs. 21, and 16 vs. 24. The correct quantities were 9, 12, 10, 18, 14, and 16, respectively. The pairs tested for ratio 4/5 were 8 vs. 10, 12 vs. 15, 16 vs. 20, 20 vs. 25, 24 vs. 30, and 28 vs. 35. The correct quantities were 8, 15, 16, 25, 24 and 35, respectively. Each pair was tested twice.

RESULTS

Descriptive Statistics

Four- to 8-year-olds' performances on the number-naming task, the verbal-counting task, non-symbolic, symbolic comparison tasks, mapping tasks, and TEMA-3 were presented in **Table 1**. One-sample *t*-tests showed that all age groups performed well above chance-level in the non-symbolic comparison task. However, only 5- to 8-year-olds performed above chance in symbolic comparison task. Six- to 8-year-olds performed above chance in mapping tasks, but not 4-to 5-year-olds.

The Development Trajectories of Non-symbolic and Symbolic Representation Abilities

Four-year-old children performed at chance level in symbolic comparison task. Therefore, their data were eliminated from the following analysis. In order to provide detailed descriptions on the development of non-symbolic and symbolic representation capacities during childhood, we conducted a 2 (Task: non-symbolic and symbolic) $\times$ 4 (Ratio: 2:3, 3:4, 4:5, 5:6) $\times$ 4 (Age: 5, 6, 7, 8 years old) repeated measures ANOVA on children's performance accuracy. Mauchly's test indicated that the assumption of sphericity had been violated for Ratio, $\chi^2(5) = 19.256$, $p = 0.002$. Therefore, we corrected the degrees of freedom by using the Greenhouse–Geisser estimates. The Box's *M* test result for the homogeneity of variance hypothesis was significant (Box's *M* test = 324.071, $F = 2.742$, $p = 0.000$). Therefore, we showed the results of Friedman and Wilcoxon non-parametric test at the same time. Results demonstrated the main effects of Ratio, $F(2.800,489.916) = 43.220$, $p < 0.001$, $\eta_p^2 = 0.198$, Task, $F(1.000,175.000) = 11.611$, $p < 0.010$, $\eta_p^2 = 0.062$, Age, $F(3,175) = 12.312$, $p < 0.001$, $\eta_p^2 = 0.174$, a significant interaction between Task and Ratio, $F(2.855,504.891) = 19.649$, $p < 0.001$, $\eta_p^2 = 0.101$, and a marginal significant interaction between Task

and Age, $F(3.000,175.000) = 2.639$, $p = 0.051$, $\eta_p^2 = 0.043$. Further simple effect analyses (and the Friedman non-parametric test) for the interaction between Task and Ratio indicated that, both in non-symbolic and symbolic comparison tasks, there was a significant ratio effect, $F_{\text{non-symbolic}}(3,525) = 17.720$, $p < 0.001$, $\eta_p^2 = 0.091$ [$\chi^2(3) = 68.208$, $p < 0.001$], $F_{\text{symbolic}}(3,525) = 43.660$, $p < 0.001$, $\eta_p^2 = 0.199$ [$\chi^2(3) = 104.614$, $p < 0.001$]. Further simple effect analyses for the interaction between Task and Age demonstrated that, 5-year-olds were better at non-symbolic task than symbolic task, $F(1,175) = 12.910$, $p < 0.001$, $\eta_p^2 = 0.068$, but other age groups performed equally on the symbolic and the non-symbolic task, $F_{6\text{-year-olds}}(1,175) = 2.190$, $p = 0.141$, $F_{7\text{-year-olds}}(1,175) = 2.500$, $p = 0.116$, $F_{8\text{-year-olds}}(1,175) = 0.010$, $p = 0.914$ (See **Figure 2**). The Wilcoxon non-parametric test confirmed the similar effect of age, $Z_{5\text{-year-olds}} = -2.570$, $p < 0.050$, *Z*s for other age groups were from -1.504 to -0.296 , *P*s > 0.050 . These results suggested the advantage of non-symbolic numerical representations over symbolic ones was salient in early childhood. However, after 5, as children learn more about the symbolic representation system, their advantage in non-symbolic representations disappeared.

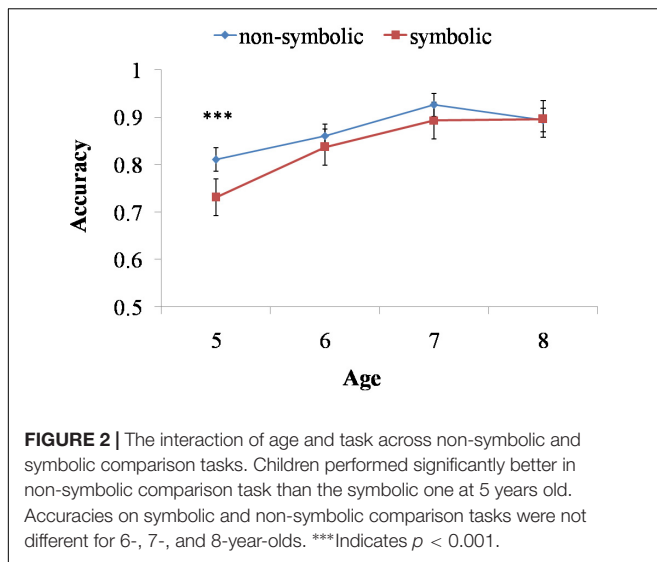
The Associations Between Numerical Representation Skills and Mathematical Ability

Correlation coefficients and partial correlation coefficients (controlling for age) between different tasks were presented **Table 2**. There were strong associations between number-naming, verbal-counting skills, non-symbolic and symbolic comparison tasks and mathematical ability, but after controlling for age, the correlations between verbal-counting abilities, numerical comparison skills, and mathematical ability were not anymore significant. This indicated that the verbal-counting ability had no significantly direct effect on non-symbolic, symbolic comparison, and mathematical skills. However, both

TABLE 1 | Children's performance in numerical comparisons, mapping tasks, and mathematical ability test.

	4 years old		5 years old		6 years old		7 years old		8 years old	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Na	1.600	1.195	3.390	1.715	4.980	0.127	5.000	0.000	5.000	0.000
VC	2.150	1.966	5.440	3.165	9.160	1.883	10.000	0.000	10.000	0.000
N	0.649***	0.158	0.811***	0.169	0.861***	0.126	0.926***	0.096	0.894***	0.133
	<i>d</i> = 0.689		<i>d</i> = 0.843		<i>d</i> = 0.944		<i>d</i> = 0.976		<i>d</i> = 0.948	
S	0.517	0.053	0.731***	0.161	0.837***	0.138	0.893***	0.098	0.896***	0.084
			<i>d</i> = 0.822		<i>d</i> = 0.925		<i>d</i> = 0.971		<i>d</i> = 0.978	
NS	0.511	0.037	0.513	0.135	0.568**	0.147	0.564**	0.141	0.562**	0.142
					<i>d</i> = 0.424		<i>d</i> = 0.418		<i>d</i> = 0.404	
SN	0.502	0.042	0.558	0.171	0.578***	0.159	0.579***	0.122	0.589***	0.150
					<i>d</i> = 0.442		<i>d</i> = 0.548		<i>d</i> = 0.513	
TEMA-3	110.77	7.316	111.43	8.449	108.75	7.534	110.36	9.923	112.53	6.541

Na, number-naming ability; VC, verbal-counting ability; N, non-symbolic comparison task; S, symbolic comparison task; NS, non-symbolic to symbolic mapping; SN, symbolic to non-symbolic mapping. One-sample *t*-tests were used to compare children's accuracies in non-symbolic, symbolic comparison tasks, mapping tasks with the chance level, ***indicates $p < 0.001$, **indicates $p < 0.01$. *d* refers to the effect size.



correlation and partial correlation analyses showed strong associations between number-naming, numerical comparison, and mathematical skills, and between the mapping skills and symbolic representation skills. These close links between each type of skills and the mathematical ability allow us to construct a structure model to better understanding of the mechanism.

We conducted structural equation modeling (SEM) analyses to examine the associations between non-symbolic, symbolic, mapping skills, and mathematical ability using Mplus Version 7. We developed one model for the developmental period from age 5–8 (Model A) and four separate models for each age groups (see Table 3, Model B was for 5-year-olds, Model

C was for 6-year-olds, Model D was for 7-year-olds, Model E was for 8-year-olds). The SEM fit indexes (Confirmatory Fit Index and Root Mean Square Error of Approximation) suggested a goodness of fit for all five models (see Table 3). Model A, capturing the entire developmental period from age 5 to 8, explained 42.1% of the variance in mathematical ability. It revealed a direct effect of symbolic skills on mapping skills and mathematical ability (see the effect values marked in Table 3). Children's non-symbolic skills affected their mathematical ability indirectly, via symbolic skills. Comparing the four models for different age groups, we found that this indirect effect of non-symbolic skills on mathematical ability was only significant for 5- and 6-year-olds, but not for 7- and 8-year-olds. The direct effect of symbolic skills on mathematical ability was significant for 5-, 6-, and 7-year-old, but not for 8-year-olds. Furthermore, the effect values of both non-symbolic and symbolic numerical representation skills on mathematical performance declined with age (see effect values marked in Table 3). Across models, we did not found significant effects of mapping skills on mathematical ability.

DISCUSSION

We investigated two issues in our study. First, we showed detailed developmental trajectories of non-symbolic and symbolic representation skills from age 4 to 8. Children were able to do non-symbolic representation task at age 4. Five-year-olds performed better in the non-symbolic task than they did in the symbolic one. However, after 5, as children learn more about the symbolic representation system, their advantage of non-symbolic skills disappeared. Second, we found a significant effect of symbolic skills on math performance and an indirect effect of non-symbolic skills on the mathematical ability via symbolic skills. Both the direct effect of symbolic skills and the indirect effect non-symbolic skills declined with age. This suggests that non-symbolic and symbolic numerical representation skills may no longer be the major factors for math performance of children in primary school.

The Developmental Trajectories of Non-symbolic and Symbolic Representation Abilities

A variety of studies suggested the inherent and universal nature of non-symbolic representation (Wynn, 1992; Pica et al., 2004; Flombaum et al., 2005). The current study demonstrated children as young as 4 years old were able to represent and compare non-symbolic quantities of range 5 to 50 successfully and flexibly. Similar paradigm was also used by Toll et al.'s (2015) testing children's non-symbolic comparison for numbers ranging from 1 to 100. Children performed well on their non-symbolic comparison task starting from age 4. For a smaller and narrower range of number from 4 to 15, researchers found similar results in 4-year-olds (Libertus et al., 2011). Wagner and Johnson (2011) assessed non-symbolic comparison skills with numbers range 1–50. They found 3-year-olds performed above

TABLE 2 | Correlation coefficients and partial correlation coefficients (controlling for age) between different numerical tasks.

		Na	VC	N	S	NS	SN
Na	r	1.000					
	r_p	1.000					
VC	r	0.867***	1.000				
	r_p	0.440***					
N	r	0.556***	0.548***	1.000			
	r_p	0.218**	0.045				
S	r	0.555***	0.251**	0.517***	1.000		
	r_p	0.435***	0.129	0.465**			
NS	r	0.055	0.077	0.088	0.155*	1.000	
	r_p	0.060	0.069	0.091	0.163*		
SN	r	0.086	−0.026	0.109	0.198**	0.069	1.000
	r_p	0.084	−0.039	0.104	0.204**	0.069	
TEMA-3	r	0.809***	0.727***	0.570***	0.568***	0.118	0.087
	r_p	0.296***	0.044	0.228**	0.426***	0.185*	0.108

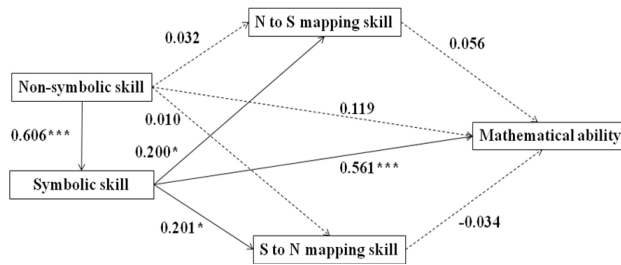
*Indicates $p < 0.050$, **indicates $p < 0.010$, ***indicates $p < 0.001$. Na, number-naming ability; VC, verbal-counting ability; N, non-symbolic comparison task; S, symbolic comparison task; NS, non-symbolic to symbolic mapping; SN, symbolic to non-symbolic mapping.

TABLE 3 | The SEM of non-symbolic, symbolic representation, mapping skill, and mathematical ability.

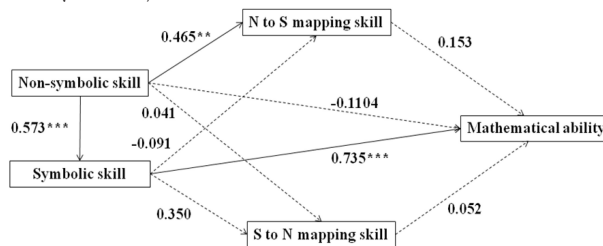
Model A (5- to 8-year-olds)

Bootstrap: $\chi^2(1) = 0.379$, $p = 0.538$ CFI = 1.000

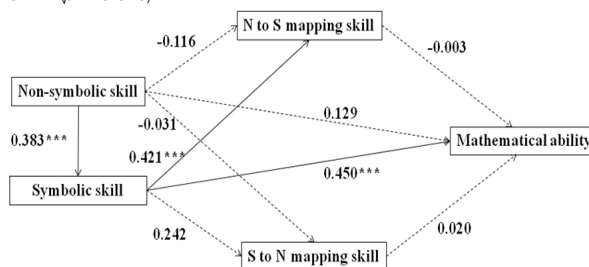
RMSEA = 0.000



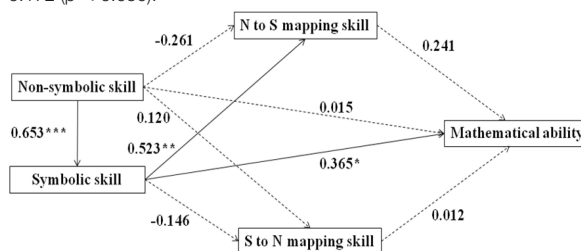
SEM analyses revealed the indirect effect value of the non-symbolic number skills on the mathematical ability is 0.340 ($p < 0.000$).

Model B (5-year-olds) Bootstrap: $\chi^2(1) = 0.130$, $p = 0.719$ CFI = 1.000 RMSEA = 0.000

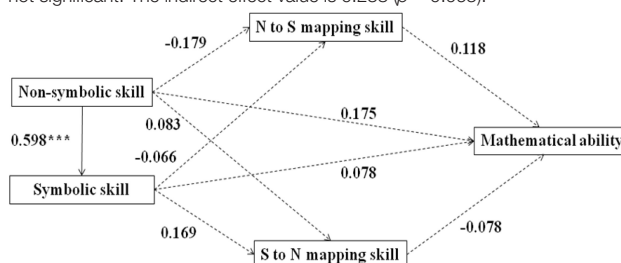
SEM analyses revealed the indirect effect value of the non-symbolic number skills on the mathematical ability is 0.421 ($p < 0.010$).

Model C (6-year-olds) Bootstrap: $\chi^2(1) = 0.083$, $p = 0.773$ CFI = 1.000 RMSEA = 0.000

SEM analyses revealed the indirect effect value of the non-symbolic number skills on the mathematical ability is 0.172 ($p < 0.050$).

Model D (7-year-olds) Bootstrap: $\chi^2(1) = 2.563$, $p = 0.109$ CFI = 0.950 RMSEA = 0.053

SEM analyses revealed the indirect effect value of the non-symbolic number skills on the mathematical ability is not significant. The indirect effect value is 0.238 ($p = 0.063$).

Model E (8-year-olds) Bootstrap: $\chi^2(1) = 0.014$, $p = 0.906$ CFI = 1.000 RMSEA = 0.000

SEM analyses revealed the indirect effect value of the non-symbolic number skills on the mathematical ability is not significant. The indirect effect value is 0.046 ($p = 0.672$).

N to S mapping skill, Non-symbolic to Symbolic mapping skill; S to N mapping skill, Symbolic to Non-symbolic mapping skill. *Indicates $p < 0.05$, **indicates $p < 0.01$, ***indicates $p < 0.001$.

chance level in non-symbolic comparison task with numerosities 1–4. To prevent children from precisely tracking dots, we used numerosities larger than 4. Although different stimuli were used in our study, the present results are still in line with previous studies, which provided evidence for the development of non-symbolic capacity after age 4 (Libertus et al., 2013; Vanbinst et al., 2015). However, for symbolic representation, our study showed 5-year-olds and older children, but not 4-year-olds, performed well in our comparison task. Similarly, previous studies (Gilmore et al., 2007; Kolkman et al., 2013) reported that children started being able to do symbolic representation task at the age of 5, before the start of formal schooling. Furthermore, researchers had found symbolic representation skills developed continuously during childhood (Li et al., 2017). These results indicate the acquired nature of the symbolic comparison skills. As a learned ability, its development is built on some more fundamental capacities, such as non-symbolic representations. Our SEM analyses showed a significant effect of non-symbolic skills on symbolic skills (see effect values in model B to model E). The indirect effect of the non-symbolic skills on mathematical abilities was carried out by symbolic skills. Therefore, we think, to some extent, the mastery of non-symbolic comparison skills was as precondition for the development of symbolic comparison skills.

There are limited studies in the field describing development trajectories of non-symbolic and symbolic comparison ability for a larger age span in childhood. Oftentimes researchers only investigated 2 to 3 age groups. For example, Xenidou-Dervou et al. (2015) focused on 5- and 6-year-olds. They also considered the developmental changes of non-symbolic and symbolic abilities. However, they used the approximate addition tasks, which were more difficult than the approximate comparison tasks in our study. In their task, children had to add the two quantities first and then to compare the numerosities, which required the arithmetic ability at the same time. Xenidou-Dervou et al. (2015) found that the ability of symbolic addition emerged around age 6. Our results provide detailed developmental trajectories of non-symbolic and symbolic comparison abilities for a larger age span in childhood. We found that 4-year-olds were able to do non-symbolic comparisons, but not symbolic comparisons. Five-year-olds were able to do both types of comparisons, but they performed better at the non-symbolic task than the symbolic one. However, this performance difference disappeared around the age of six. We think these developmental changes may be related to the different characteristics of non-symbolic and symbolic skills. Non-symbolic representation ability is inherent, shared by humans and animals (Wynn, 1992; Pica et al., 2004; Flombaum et al., 2005). However, symbolic comparison ability is affected by education (Xenidou-Dervou et al., 2015), and its emergence requires a certain foundation (Kolkman et al., 2013). Many researchers have found that children's symbolic representation skill will rapidly increase in the 1st grade (Xenidou-Dervou et al., 2015; Li et al., 2017). Therefore, we observed that children could pass non-symbolic tasks at a very young age, but they were not able to pass symbolic representation tasks until 5 years old. However, with more education, children's symbolic skills

improve rapidly and their advantage in non-symbolic skills disappears around 6 years old.

The Associations Between Numerical Representation Skills and Mathematical Ability

Fazio et al. (2014) proposed three hypotheses about the relationship between non-symbolic, symbolic skills, and mathematical ability: (1) non-symbolic skills have indirect effects on mathematics achievement. That is, children with better non-symbolic skills acquire the symbolic numerical system more easily, which in turn improves their mathematical ability; (2) non-symbolic skills have both direct and indirect effects on mathematics achievement; (3) non-symbolic and symbolic skills may independently affect overall mathematics achievement. In the current study, we found an indirect effect of non-symbolic skills on mathematical abilities via symbolic skills, which supports Fazio et al.'s (2014) first hypothesis. Similar results were also found by van Marle et al. (2014), who assessed non-symbolic, symbolic skills, and mathematics achievement in 4-year-olds and found that the relation between non-symbolic skills and mathematics achievement was fully mediated by children's symbolic skills. Differently, a significant positive correlation between the precision of non-symbol quantity and mathematical achievement in 3- to 5-year-old children was reported by Libertus et al. (2012). They used children's ANS acuity, rather than accuracy, as an indicator of children's non-symbolic skill. The ANS acuity is represented by Weber's fraction, which is derived from the theoretical hypothesis of psychophysics. It is an indirect indicator for numerical representation ability. However, the ANS accuracy illustrates numerical representation ability more directly. This measurement difference might result the different findings here. On the other hand, as shown in previous studies (Kolkman et al., 2013; Toll et al., 2015), we also found a significant effect of symbolic skills on mathematical ability.

In addition, we found that children's mapping ability had no significant effects on their mathematical ability. However, using similar paradigm, Mundy and Gilmore's (2009) found children's bi-directional mapping ability predicted their mathematical achievement significantly. This result might be because, comparing to our tasks using comparison ratios of 2/3 and 4/5, Mundy and Gilmore's (2009) tasks were easier. They used relative easy comparison ratios of 1/2 and 2/3. Other researchers used different paradigms to assess children's mapping ability. For example, Kolkman et al. (2013) found mapping skills was an important predictor for math performance. However, they used symbolic number-lines and symbolic comparison tasks, which are very different from our bi-directional mapping task. Therefore different results were generated.

Finally, we found the associations between numerical representation skills and mathematical abilities varied across age groups. The indirect effect of non-symbolic skills on mathematical abilities was only significant for 5- and 6-year-olds, but not for 7- and 8-year-olds. The direct effect of symbolic skills on mathematical abilities was significant for 5-, 6-, and 7-year-olds, but not for 8-year-olds. In general, the impacts of

non-symbolic and symbolic numerical representation skills on mathematical performance both declined with age. We think the result may suggest, with age, non-symbolic, and symbolic numerical representation skills were no longer major factors for math performance. Similar developmental trend has been found in previous studies as well. A significant positive correlation between the non-symbolic skill and mathematical achievement was reported for 3- to 5-year-olds (Libertus et al., 2012); with age, this positive correlation disappeared for 6- to 8-year-olds (Holloway and Ansari, 2009). Meanwhile, there are studies (Halberda et al., 2008; Bugden and Ansari, 2011; Linsen et al., 2015) showed correlations between numerical representation skills and mathematical ability throughout childhood. However, their methods were quite different from ours. For example, instead of TEMA-3, Bugden and Ansari (2011) used two mathematics subtests from the Woodcock Johnson III and Linsen et al. (2015) used multi-digit subtraction task to assess children's mathematical ability. Also, many of previous studies only investigated 2 to 3 age groups, which may affect how their results can be generalized.

Limitations and Future Research

The current study has limitations and therefore requests future research to further clarify these questions. First, with the cross-sectional design of the current study, the developmental information provided by the data was limited. We were not able to examine longitudinally interactions of non-symbolic and symbolic representation skills and their association with mathematical ability. This requests future research to clarify the issue. In fact, we are currently working on the follow-up of this

study. With the longitudinal data, we would be able to draw a more comprehensive picture on the development of children's numerical representation capacities and their association with mathematical performance. Second, in this study, we only considered numerosities larger than 4, which made tasks difficult for 4-year-olds. The reason we used numerosities larger than 4 is to prevent children from precisely tracking dots, because previous research (Feigenson et al., 2004) shown that children developed a system to keep track of small numbers precisely from very young. However, with numerosities smaller than 4, we may be able to capture 4-year-olds' performance in the symbolic comparison task. Future research needs to address this issue and compare children's non-symbolic and symbolic comparison skills and mapping ability for both large and small numerosities.

AUTHOR CONTRIBUTIONS

YL contributed to conception and design, on acquisition and interpretation of data, and on drafting of manuscript. YC contributed to conception and design and on interpretation of data. MZ contributed to drafting of manuscript. XZ and ZD contributed on interpretation of data. SY contributed to making the experimental materials and entering the data.

FUNDING

This work was supported by grants from National Natural Science Foundation of China (31271106) and National Social Science Foundation of China (14ZDB160) to YC.

REFERENCES

- Barth, H., Beckmann, L., and Spelke, E. S. (2008). Nonsymbolic, approximate arithmetic in children: abstract addition prior to instruction. *Dev. Psychol.* 44, 1466–1477. doi: 10.1037/a0013046
- Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., and Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition* 98, 199–222. doi: 10.1016/j.cognition.2004.09.011
- Barth, H., La Mont, K., Lipton, J., and Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14116–14121. doi: 10.1073/pnas.0505512102
- Berkowitz, T., Schaeffer, M. W., Schaeffer, E. A., Peterson, L., Gregor, C., Levine, S. C., et al. (2015). Math at home adds up to achievement in school. *Science* 350, 196–198. doi: 10.1126/science.aac7427
- Bonny, J. W., and Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: evidence from the preschool years. *J. Exp. Child Psychol.* 114, 375–388. doi: 10.1016/j.jecp.2012.09.015
- Brankaer, C., Ghesquiere, P., and De Smedt, B. (2014). Children's mapping between Non-symbolic and symbolic numerical magnitudes and its association with timed and untimed tests of mathematics achievement. *PLoS One* 9:e93565. doi: 10.1371/journal.pone.0093565
- Bugden, S., and Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition* 118, 32–44. doi: 10.1016/j.cognition.2010.09.005
- Castronovo, J., and Goebel, S. M. (2012). Impact of high mathematics education on the number sense. *PLoS One* 7:e33832. doi: 10.1371/journal.pone.0033832
- Chen, Q., and Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychol.* 148, 163–172. doi: 10.1016/j.actpsy.2014.01.016
- Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics, Revised and Updated Edition*. New York, NY: Oxford University Press, 352.
- DeWind, N. K., and Brannon, E. M. (2012). Malleability of the approximate number system: effects of feedback and training. *Front. Hum. Neurosci.* 6:68. doi: 10.3389/fnhum.2012.00068
- Fazio, L. K., Bailey, D. H., Thompson, C. A., and Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *J. Exp. Child Psychol.* 123, 53–72. doi: 10.1016/j.jecp.2014.01.013
- Feigenson, L., Carey, S., and Spelke, E. S. (2002). Infants' discrimination of number vs. continuous extent. *Cogn. Psychol.* 44, 33–66. doi: 10.1006/cogp.2001.0760
- Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307–314. doi: 10.1016/j.tics.2004.05.002
- Flombaum, J. I., Junge, J. A., and Hauser, M. D. (2005). Rhesus monkeys (*Macaca mulatta*) spontaneously compute addition operations over large numbers. *Cognition* 97, 315–325. doi: 10.1016/j.cognition.2004.09.004
- Friso-van Den Bos, I., Kroesbergen, E. H., Van Luit, J. E. H., Xenidou-Dervou, I., Jonkman, L. M., Van der Schoot, M., et al. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *J. Exp. Child Psychol.* 134, 12–29. doi: 10.1016/j.jecp.2015.02.002
- Gilmore, C. K., McCarthy, S. E., and Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature* 447, 589–591. doi: 10.1038/nature05850
- Ginsburg, H. P., and Baroody, A. J. (2003). *Test of Early Mathematics Ability*, 3rd Edn. Austin, TX: Pro-Ed.
- Gobel, S. M., Watson, S. E., Lervag, A., and Hulme, C. (2014). Children's arithmetic development: it is number knowledge, not the approximate number sense, that counts. *Psychol. Sci.* 25, 789–798. doi: 10.1177/0956797613516471

- Halberda, J., Mazocco, M. M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* 455, 665–668. doi: 10.1038/nature07246
- Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children's mathematics achievement. *J. Exp. Child Psychol.* 103, 17–29. doi: 10.1016/j.jecp.2008.04.001
- Hyde, D. C., Khanum, S., and Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition* 131, 92–107. doi: 10.1016/j.cognition.2013.12.007
- Inglis, M., Attridge, N., Batchelor, S., and Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: but only in children. *Psychon. Bull.* 18, 1222–1229. doi: 10.3758/s13423-011-0154-1
- Izard, V., and Dehaene, S. (2008). Calibrating the mental number line. *Cognition* 106, 1221–1247. doi: 10.1016/j.cognition.2007.06.004
- Kolkman, M. E., Kroesbergen, E. H., and Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learn. Instr.* 25, 95–103. doi: 10.1016/j.learninstruc.2012.12.001
- Li, Y., Zhang, M., Chen, Y., Zhu, X., Deng, Z., and Yan, S. (2017). Children's Non-symbolic, symbolic addition and their mapping capacity at 4–7 years old. *Front. Psychol.* 8:1203. doi: 10.3389/fpsyg.2017.01203
- Libertus, M. E., Feigenson, L., and Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Dev. Sci.* 14, 1292–1300. doi: 10.1111/j.1467-7687.2011.01080.x
- Libertus, M. E., Feigenson, L., and Halberda, J. (2013). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *J. Exp. Child Psychol.* 116, 829–838. doi: 10.1016/j.jecp.2013.08.003
- Libertus, M. E., Odic, D., and Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychol.* 141, 373–379. doi: 10.1016/j.actpsy.2012.09.009
- Linsen, S., Verschaffel, L., Reynvoet, B., and De Smedt, B. (2015). The association between numerical magnitude processing and mental versus algorithmic multi-digit subtraction in children. *Learn. Instr.* 35, 42–50. doi: 10.1016/j.learninstruc.2014.09.003
- Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS One* 6:e23749. doi: 10.1371/journal.pone.0023749
- Mundy, E., and Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *J. Exp. Child Psychol.* 103, 490–502. doi: 10.1016/j.jecp.2009.02.003
- Mussolin, C., Nys, J., Content, A., and Leybaert, J. (2014). Symbolic number abilities predict later approximate number system acuity in preschool children. *PLoS One* 9:e91839. doi: 10.1371/journal.pone.0091839
- Newcomer, P. (2001). *Diagnostic Achievement Battery*, 3rd Edn. Austin, TX: Pro Ed.
- Pica, P., Lemer, C., Izard, W., and Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306, 499–503. doi: 10.1126/science.1102085
- Praet, M., and Desoete, A. (2014). Number line estimation from kindergarten to grade 2: a longitudinal study. *Learn. Instr.* 33, 19–28. doi: 10.1016/j.learninstruc.2014.02.003
- Price, G. R., Palmer, D., Battista, C., and Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychol.* 140, 50–57. doi: 10.1016/j.actpsy.2012.02.008
- Rousselle, L., Palmers, E., and Noel, M.-P. (2004). Magnitude comparison in preschoolers: what counts. Influence of perceptual variables. *J. Exp. Child Psychol.* 87, 57–84. doi: 10.1016/j.jecp.2003.10.005
- Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: what underlies mathematics achievement? *J. Exp. Child Psychol.* 114, 418–431. doi: 10.1016/j.jecp.2012.10.012
- Skwarchuk, S., Sowinski, C., and LeFevre, J. (2014). Formal and informal home learning activities in relation to children's early numeracy and literacy skills: the development of a home numeracy model. *J. Exp. Child Psychol.* 121, 63–84. doi: 10.1016/j.jecp.2013.11.006
- Toll, S. W. M., Van Viersen, S., Kroesbergen, E. H., and Van Luit, J. E. H. (2015). The development of (non-)symbolic comparison skills throughout kindergarten and their relations with basic mathematical skills. *Learn. Individ. Differ.* 38, 10–17. doi: 10.1016/j.lindif.2014.12.006
- van Marle, K., Chu, F. W., Li, Y., and Geary, D. C. (2014). Acuity of the approximate number system and preschoolers' quantitative development. *Dev. Sci.* 17, 492–505. doi: 10.1111/desc.12143
- Vanbinst, K., Ghesquière, P., and De Smedt, B. (2015). Does numerical processing uniquely predict first graders' future development of single-digit arithmetic? *Learn. Individ. Differ.* 37, 153–160. doi: 10.1016/j.lindif.2014.12.004
- Wagner, J. B., and Johnson, S. C. (2011). An association between understanding cardinality and analog magnitude representations in preschoolers. *Cognition* 119, 10–22. doi: 10.1016/j.cognition.2010.11.014
- Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Rolling Meadows, IL: Riverside Publishing.
- Wynn, K. (1992). Addition and subtraction by human infants. *Nature* 358, 749–750. doi: 10.1038/358749a0
- Xenidou-Dervou, I., Gilmore, C., van der Schoot, M., and van Lieshout, E. C. D. M. (2015). The developmental onset of symbolic approximation: beyond nonsymbolic representations, the language of numbers matters. *Front. Psychol.* 6:487. doi: 10.3389/fpsyg.2015.00487

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

Copyright © 2018 Li, Zhang, Chen, Deng, Zhu and Yan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Taking Language out of the Equation: The Assessment of Basic Math Competence Without Language

Max Greisen^{1*}, Caroline Hornung², Tanja G. Baudson¹, Claire Muller², Romain Martin² and Christine Schiltz¹

¹ Cognitive Science and Assessment Institute, University of Luxembourg, Luxembourg, Luxembourg, ² Luxembourg Centre for Educational Testing, University of Luxembourg, Luxembourg, Luxembourg

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

Delphine Sasanguie,
KU Leuven Kulak, Belgium
Robert Reeve,
University of Melbourne, Australia

*Correspondence:

Max Greisen
max.greisen@uni.lu

Specialty section:

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

Received: 23 February 2018

Accepted: 07 June 2018

Published: 26 June 2018

Citation:

Greisen M, Hornung C, Baudson TG,
Muller C, Martin R and Schiltz C
(2018) Taking Language out of the
Equation: The Assessment of Basic
Math Competence Without Language.
Front. Psychol. 9:1076.
doi: 10.3389/fpsyg.2018.01076

While numerical skills are fundamental in modern societies, some estimated 5–7% of children suffer from mathematical learning difficulties (MLD) that need to be assessed early to ensure successful remediation. Universally employable diagnostic tools are yet lacking, as current test batteries for basic mathematics assessment are based on verbal instructions. However, prior research has shown that performance in mathematics assessment is often dependent on the testee's proficiency in the language of instruction which might lead to unfair bias in test scores. Furthermore, language-dependent assessment tools produce results that are not easily comparable across countries. Here we present results of a study that aims to develop tasks allowing to test for basic math competence without relying on verbal instructions or task content. We implemented video and animation-based task instructions on touchscreen devices that require no verbal explanation. We administered these experimental tasks to two samples of children attending the first grade of primary school. One group completed the tasks with verbal instructions while another group received video instructions showing a person successfully completing the task. We assessed task comprehension and usability aspects both directly and indirectly. Our results suggest that the non-verbal instructions were generally well understood as the absence of explicit verbal instructions did not influence task performance. Thus we found that it is possible to assess basic math competence without verbal instructions. It also appeared that in some cases a single word in a verbal instruction can lead to the failure of a task that is successfully completed with non-verbal instruction. However, special care must be taken during task design because on rare occasions non-verbal video instructions fail to convey task instructions as clearly as spoken language and thus the latter do not provide a panacea to non-verbal assessment. Nevertheless, our findings provide an encouraging proof of concept for the further development of non-verbal assessment tools for basic math competence.

Keywords: nonverbal, assessment, mathematics, language, dyscalculia, video, instruction, screener

INTRODUCTION

Basic counting and arithmetic skills are necessary to manage many aspects of life. Although primary education focuses on these subjects, 5–7% of the general population suffer from mathematical learning difficulties (MLD) (Butterworth et al., 2011), often leading to dependence on other people or technology.

Early diagnostic is key to remedying MLD (Gersten et al., 2005). Basic mathematical skills, e.g., counting, quantity comparison, ordering, and simple arithmetic are the strongest domain-specific predictors for mathematical performance in later life (Desoete et al., 2009; Jordan et al., 2010; LeFevre et al., 2010; Hornung et al., 2014). Valid MLD assessments exist in various forms and for all ages (van Luit et al., 2001; Haffner et al., 2005; Schaupp et al., 2007; Noël et al., 2008; Aster et al., 2009; Ricken et al., 2011). However, all of them rely on verbal instructions and (in part) verbal tasks.

This is a problem. First, performance in mathematical tests is predicted by the pupils' proficiency in the instruction language (Abedi and Lord, 2001; Hickendorff, 2013; Paetsch et al., 2016). Others have shown that the complexity of mathematical language content of items is predictive of performance (Haag et al., 2013; Purpura and Reid, 2016). Diagnostic tools for MLD relying on language may therefore significantly bias performance in test-takers that are not proficient in the test language, leading to invalid results (see Scarr-Salapatek, 1971; Ortiz and Dynda, 2005 for similar considerations concerning intelligence testing). Furthermore, the match between math learners' language profiles and the linguistic context in which mathematical learning takes place plays a critical role in the acquisition and use of basic number knowledge. Matching language contexts improve bilinguals' arithmetic performance in their second language (Van Rinsveld et al., 2016), and neural activation patterns of bilinguals solving additions differ depending on the language they used, suggesting different problem-solving processes (Van Rinsveld et al., 2017).

In linguistically homogeneous societies, where the mother tongue of most primary school children matches the language of instruction and assessment tools, this is less of a problem. It is however critical in societies with high immigration and, therefore, linguistically diverse primary school populations. In Luxembourg, for instance, where the present project is located, currently 62% of the primary school students are not native Luxembourgish speakers (Ministère de l'éducation nationale de l'enfance et de la Jeunesse, 2015). Due to migration, multilingual classrooms are steadily becoming the rule rather than the exception (e.g., from 42% foreign speakers in 2004 to 62% in 2014) (Ministère de l'éducation nationale de l'enfance et de la Jeunesse, 2015), likely increasing the urgency of the problem in the future.

Even in traditionally multilingual contexts, diagnostic tools for the assessment of basic numerical abilities in early childhood are available in a few selected languages only, usually those that are best understood by most, yet not necessarily all students. As described above, this leads to invalid conclusions about non-native speakers' ability. In addition, comparisons between

different tools and even different linguistic versions of the same tool are difficult because the norms they are based on are usually collected in linguistically homogenous populations and can thus not be extrapolated to populations with different linguistic profiles.

The present study originated in a project that aims to develop a test of basic numerical competencies which circumvents linguistic interference by relying on non-verbal instructions and task content. In the field of intelligence assessment, the acknowledgment of language interference has led to the development of numerous non-verbal test batteries (Cattell and Cattell, 1973; Lohman and Hagen, 2001; Naglieri, 2003; Feis, 2010). However, these tools tackle only the problem of verbal tasks, not of verbal instructions. The same is true for numeracy assessment. Although many test batteries (e.g., Tedi-MATH, Zareki-R, ERT0+, OTZ, Marko-D, to name a few) use non-verbal and non-symbolic tasks (e.g., arithmetic, counting, or logical operations on numbers), they still rely on verbal instructions, which may limit the testee's access to the content. Linguistic simplification of mathematics items can improve performance for language minority students (Haag et al., 2014). However, we think that for many simple tasks, verbal content and instructions can be avoided altogether. These tasks that children of (above-) average ability usually solve easily are crucial to the diagnosis of MLD, as they allow for a differentiation of children's numerical abilities at the bottom end of the ability distribution. Hence, non-verbal assessment of basic mathematical skills may help identify children in need of intervention at an early age and independently of their linguistic abilities, thus reducing the bias that common assessments often suffer from. Comparable approaches have been taken in the field of intelligence testing for the hearing-impaired, in which pantomime instructions for the Wechsler performance scale have been explored (Courtney et al., 1984; Braden and Hannah, 1998).

With this goal in mind, using available test batteries and the official study plan (MENFP, 2011) as a reference for task content and design, we developed different task types for which a valid non-verbal computerized implementation was possible. Governmental learning goals for preschool mathematics include but are not limited to: Ability to represent numbers with concrete material, ordering abilities (range 0–10), definition, resolution & interpretation of an arithmetical (addition/subtraction) problem based on images and mental addition/subtraction (range 0–10).

The tasks we developed encompass and measure all the above competencies: Quantity representation, ordering abilities as well as symbolic and non-symbolic arithmetic. We chose to add a quantity comparison task as it has been found to be one of the most consistent predictors of later math performance (e.g., De Smedt et al., 2009; Sasanguie et al., 2012; Nosworthy et al., 2013; Brankaer et al., 2017; see Schneider et al., 2017 for a meta-analysis). Instead of using verbal instructions, we convey task requirements with the use of videos that show successful task completion and interactions with the tasks from a first-person point of view. Prior research has shown improved performance in a computerized number-line estimation task for participants who viewed videos of a model participant's eye gaze or mouse movements, compared to control conditions

both with and without anchor points (Gallagher-Mitchell et al., 2017).

The aims of the present study were to evaluate whether basic math competence can be assessed on a tablet PC without language instructions and whether the mode of instruction affects performance. To this end, we designed a set of computerized tasks based on validated assessments measuring basic non-symbolic and symbolic mathematical abilities, which were administered either non-verbally (using computer-based demonstrations; experimental condition) or traditionally (using verbal instructions; control condition). Because young school children's attention span is limited (Pellegrini and Bohn, 2005), some of the tasks were administered to one sample (Sample 1) in a first study and the remainder to another sample (Sample 2) in a second study 5 months later. First, considering that the non-verbal mode of instruction was new, we examined possible difficulties both directly (understanding of feedback and navigation) and indirectly (repeated practice sessions). Second, though tasks were derived from field-tested assessments, performance on the new tasks was correlated with performance on two standardized and one self-developed measure in order to ensure task validity. Third, we examined students' performance compared by condition and overall. Considering the novelty of the non-verbal task administration, we did not specify directed hypotheses but examined this question exploratively.

METHODS

Participants

Table 1 shows participant demographics, language background and socio-economic status. The ISEI is the International Socio-Economic Index of Occupational Status, used in large scale assessments. It ranges from 16 (e.g., agricultural worker) to 90 (e.g., judge). An average ISEI of 50 will thus indicate above average socio-economic status. As we could not directly assess socio-economic status in our studies, ISEI was estimated based on the communes in which the studies took place. This data is publicly available and in Luxembourg the communes average ISEI ranges from 35 to 65. All participants were recruited from first grade in Luxembourg's primary schools with the authorization of the Ministry of Education and the directors of the participating school sectors. Participants from the first sample were tested after 5 weeks of schooling while participants from the second sample were tested after 28 weeks of schooling. Teachers interested to participate in the study with their classes received information and consent letters for the pupil's legal representatives. Only pupils whose parents consent was obtained participated in this study. All children in Luxembourg spend two obligatory years in preschool and about a third of them participate in an optional third year of preschool prior to the two mandatory years (Lenz, 2015).

Materials

Experimental Tasks

As mentioned, the two samples received different types of tasks. In the following, all task types will be described in order of their administration. The number in parentheses after each task name

indicates the sample it was administered to. Example images for each task are presented in Figure 1.

Quantity Correspondence (S1)

The first task required determination of the exact quantity of the target display and choosing the response display with the corresponding quantity (both ranging from 1 to 9). Each item consisted of a target quantity displayed at the center of the screen (stimulus). The nature of the quantity was varied and was either non-symbolic (based on real objects [fruit], abstract [dot collections]) or symbolic (Arab numerals). In the lower part of the screen, three different quantities were displayed to the participant from which he/she was to choose the one corresponding to the stimulus (multiple-choice images). The item pool consisted of five subgroups of items containing four items each:

1. Non-symbolic, *identical* objects for stimulus and multiple-choice images
2. Non-symbolic, *different* objects for stimulus and multiple-choice images
3. Non-symbolic, collections of black dots of variable sizes and configurations
4. Symbolic, Arabic numerals in both stimulus and multiple-choice images
5. Mixed (combinations of the preceding characteristics)

Image characteristics (object area, total occupied area, etc.) were manually randomized but not systematically controlled for.

Quantity Comparison (S1)

The second task required determining and choosing the larger of two quantities (range: 1–9) displayed at the center of the screen. The nature of the quantities was varied similarly to the first task:

1. Non-symbolic, each quantity being composed of different objects (4 items)
2. Non-symbolic, each quantity being composed of collections of black dots of variable sizes and configurations (4 items)
3. Symbolic, at least one of the two displays showing an Arabic numeral (4 items)

Ordering (S1)

The third task required reordering 4 images by increasing quantities (range 1–9). The characteristics were divided into 2 subgroups, represented by 4 items each:

1. Ordering based on non-symbolic quantity
2. Ordering based on numerical symbols (Arabic digits)

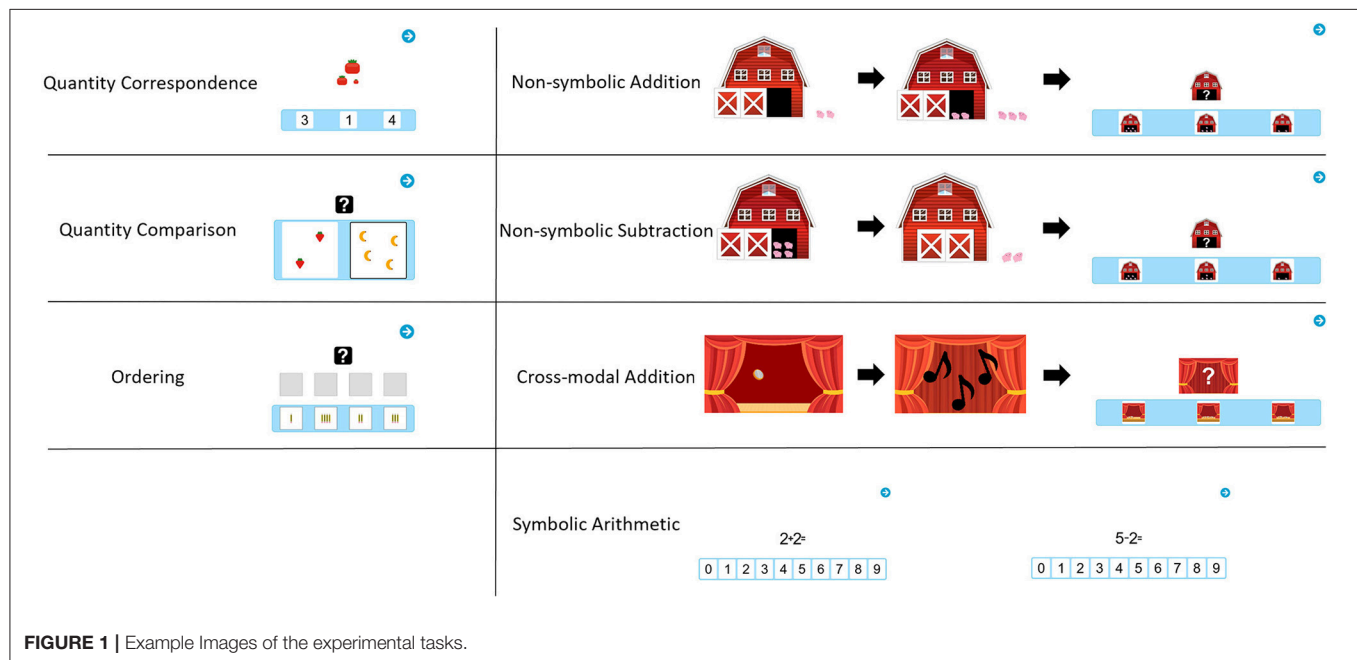
Non-symbolic Addition (S2)

The first task required to solve a non-symbolic addition problem. Participants saw an animation of 1–5 pigs entering a barn. The barn door closed. Then, the door opened again, and 1–5 more pigs entered the barn. The door closed again. The result range included the numbers from 3 to 8 only. In the **non-symbolic answer** version of this task (3 items), participants were then presented with three images containing an open barn with pigs

TABLE 1 | Participant demographics, language background and SES.

Sample	N	% girls	Age	Schooling	Language			ISEI
			M (SD)		% RO	% LG	% OT	M (SD)
Sample 1	96	53.1	6 years; 7 months (4 months)	Grade 1 (5 weeks)	30.2	55.2	14.6	50 (6.3)
Sample 2	141	48.2	7 years; 2 months (4 months)	Grade 1 (28 weeks)	55.3	34.8	9.9	47.9 (7.2)

% RO, percentage of Romance language speaking children (French, Portuguese, Italian, Spanish). % LG, percentage of children speaking Luxembourgish or German. % OT, percentage of children with other language backgrounds (Slavic, English). ISEI, International Socio-Economic Index of Occupational Status.

**FIGURE 1** | Example Images of the experimental tasks.

inside. Their task was to choose the image showing the total number of pigs left in the barn. In the **symbolic answer** version of the task (3 items), participants selected the correct number of pigs from an array of numerals from 0 to 9 in ascending order to choose from.

Non-symbolic Subtraction (S2)

The second task required solving a non-symbolic subtraction problem using the same pigs-and-barn setting described above. Participants were shown an animation of an open barn containing some pigs, after which some pigs left and the barn door closed. The minimum number of pigs displayed in a group was 2, the maximum was 9. The result range was from 1 to 6. **Symbolic** and **non-symbolic** answer versions (3 items each) were the same as above.

Crossmodal Addition (S2)

The third task for Sample 2 required solving a crossmodal addition problem using visual and auditory stimuli. Participants saw an animation of coins dropping on the floor, each one making a distinctive sound. A curtain was then closed in front of the coins. More coins dropped, but the curtain remained closed. Participants could only hear but not see the second set of coins falling. Their task was to choose the total amount of coins on the floor, both the ones they saw and heard and the ones they

only heard but did not see falling. The minimum number of coins displayed/heard was 1, the maximum was 5. The result range was from 3 to 7. In the **non-symbolic answer** version of this task (3 items), participants were presented with three images showing coins on the floor with an open curtain. Their task was to choose the image showing the total number of coins that are now on the floor. In the **symbolic answer** version of the task (3 items), participants were presented with an array of numerals from 0 to 9 in ascending order to choose from.

This task aimed to assess numerical processing at a crossmodal level, requiring a higher level of abstraction than unimodal tasks like the non-symbolic addition and subtraction tasks where only visual information is processed before answering the question. The addition of discrete sounds as stimuli adds a layer of abstraction that is not present in the other addition tasks (symbolic or non-symbolic) and ensures that responses must be based on a *truly abstract number sense, capable of representing any set of discrete elements* (Barth et al., 2003), independently from its physical nature and prior cultural learning of number symbols.

Symbolic Arithmetic: Addition and Subtraction (S2)

In this task, participants had to solve traditional symbolic arithmetic problems in the range of 0–9, both addition (6 items) and subtractions (6 items), shown at the center of the screen. The answer format in this task was symbolic only, i.e., participants

were presented with an array of numerals from 0 to 9 in ascending order below the problem to choose their answer from.

Observation and Interview Sheets

To examine the usability of instructions and task presentation, test administrators collected information about participants' behavior during testing through semi-structured observation and interview sheets. Of special interest were the observations about the general use of the tablet and the tool's navigational features as well as participants' understanding of both video and verbal instructions and feedback elements in both groups.

The following questions (yes-no format) were answered for each participant and task: (1) Did the participant understand the purpose of the smiley? (2) Did the participant understand the use of the blue arrow as a navigational tool? To this aim, the test administrators asked the participants to describe the task, the role of the smiley, and the role of the arrow and evaluated that answer as a "Yes" or a "No." These questions were followed by empty space for comments.

Demographics and Criterion Validation Tasks

After completion of the digitally administered tasks, all children received a paper notebook containing a demographic questionnaire as well as some control tasks. The questionnaire collected basic demographic data (age, gender, language spoken with mother). Control tasks were included to examine the criterion validity of the experimental tasks and were administered to both samples. The paper pencil control tasks were:

- **TTR (*Tempo Test Rekenen*)** (De Vos, 1992): a classical standardized measure of speeded arithmetic performance. Participants had 60 s for each subtest. Arithmetic difficulty increased systematically within each subtest list, with operands and results in the range of 1–100. As multiplication and division were not part of the participant's curriculum at that age, we used the addition and subtraction subtests only.
- **"How many animals?" (Counting and transcoding)**: Since all of our experimental task assume basic counting skills, we included this self-developed counting task, in which ten paper sheets displaying a randomly arranged variable number of animals (range: 3–19) were presented successively to the participants, who reported how many animals they saw. Their oral answer was noted on a coding sheet by the test administrators. Furthermore, participants wrote down their answer on a separate coding sheet included in the participant notebook. This resulted in two separate measures: one for counting (oral) and one for transcoding ability (written).
- **SYMP (*Symbolic magnitude processing test*)** (Brankaer et al., 2017): a standardized measure of symbolic number comparison performance (1- and 2-digit, ranging from 1 to 10 and from 12 to 99, respectively). It includes a motor speed control task requiring participants to cross out the black shape in pairs of black/white shapes. Participants had 30 s for each subtest. Although number comparison abilities assessed by the SYMP test do not strictly constitute a measure of curricular

learning goals, we choose to include it due to its well-recognized power to predict later differences in standardized mathematical tests and distinguish children with MLD from typically developing peers (see Schneider et al., 2017 for a meta-analysis). In contrast to the TTR scales and the counting task, correlation with the SYMP does not inform on the ability of our tasks to predict children's achievement on higher level learning goals but allows to compare performance in our tasks to another low-level predictor of later math competence.

Design and Procedure

Experimental Design

To evaluate comprehensibility and effectiveness of the video instructions in comparison to classical verbal instructions, we implemented a between-group design in the two samples. All children solved the tasks on tablet computers, but under two different conditions. In the experimental condition (non-verbal condition), instructions were conveyed through a video of a person performing specific basic mathematical tasks, followed by a green smiley indicating successful solution of the task. Importantly, children did not receive any verbal instructions in the experimental condition. In the control condition (verbal condition), children received verbal instructions in German, the official instruction language for Mathematics in elementary schools in Luxembourg. Analogous to usual classroom conditions, test administrators read the instructions aloud to the children. In both conditions, tasks were presented visually on tablet computers, either through static images or animated "short stories." In both samples, one group was allocated to the experimental non-verbal condition without language instructions and the other group was assigned to the verbal condition, respectively.

Task Presentation

The three main tasks for Sample 1 were presented on iPads using a borderless browser window. Two children were tested simultaneously. They were connected to a local server through a secured wireless network set up by the research team at each school to store and retrieve data. The tasks were implemented using proprietary web-based assessment-building software under development by the Luxembourg Centre for Educational Testing. Sample 2 worked on Chromebooks instead of iPads. The advantage of Chromebooks is that they are relatively inexpensive, are optimized for web applications, and provide both touchscreen interactivity and a physical keyboard when necessary. Four children were tested simultaneously to speed up data collection.

After the initial setup of the hardware (server, wireless connection), participants were called into the test room in groups of two (Sample 1) or four (Sample 2) and seated individually on opposite sides of the room, allowing to run multiple test sessions simultaneously. Participants were randomly assigned to one of two groups. A trained test administrator supervised each participant during the test session. Since the tasks for Sample 2 used audio material, participants were provided with headphones, which they wore during the video instructions and the tasks.

Both samples were presented with either non-verbal or verbal instructions. In the non-verbal condition (experimental group), each participant was shown three items, with the exception of the comparison task, where ten instruction items were given to account for the less salient nature of the implicit “Where is more?” instruction. The video also clarified how to proceed to the next item by the person touching a blue arrow pointing rightwards on the top right corner of the screen, after which a new item was loaded. In the verbal condition (control group), the test administrator read the standardized oral instructions to the participant in German, thus mimicking traditional teaching and test situations. The instruction was repeated by the test administrator while the first practice item was displayed to facilitate the hands-on understanding of the task. After the instruction, participants were given three practice items with the same smiley-type feedback they had just witnessed (a happy green face for correct answers, an unhappy red face for wrong answers). After successful completion of the three practice items, the application moved on to the test items. If one or more answers were wrong, all three practice items were repeated once, including those that had been solved correctly in the first trial. At the end of this second run, the application moved on to the test items, even if one or more practice items had still been answered incorrectly. After each practice session, an animation showing a traffic light switching from red to green was displayed to notify children that the test was about to start.

At the end of the three tasks, a smiley face was displayed thanking the participants for their efforts. At the end of the individual testing sessions, all participants were regrouped in their classroom to complete the pen-and-paper measures instructed orally by the test administrators.

Scoring

Scores from symbolic and non-symbolic subgroups of items in most *experimental tasks* were averaged and operationalized as

POMP (percentage of maximum performance) scores (Cohen et al., 1999), giving rise to two scores in each task. The exception was the symbolic arithmetic task in Sample 2, which by its nature included only symbolic answer formats, but offered both addition and subtraction items, producing one score for each operation type. All scores from the *criterion validation tasks* are expressed as POMP scores.

RESULTS

In line with our research questions outlined in the introduction, we will first report findings on participants’ difficulties by experimental condition, as usability represents an important prerequisite. Results on the directly assessed difficulties will focus on understanding of feedback and navigation, whereas indirectly assessed difficulties comprise findings on repeated practice. This is followed by descriptive analyses including scale quality, tests of normality, and scale intercorrelations. As we also examined the convergent validity of our tasks (another prerequisite), which were based on existing measures, we subsequently report findings on the correlations with the external measures, i.e., the paper pencil tests (see Materials section). Finally, we will compare performance by experimental condition.

Observation Data

Directly Assessed Difficulties: Understanding of Feedback and Navigation

The following results are based on the observation sheets for each task. **Table 2** shows the number of participants that understood the smiley as a feedback symbol and the number of participants that understood the arrow as a navigational interface element. Discrepancies in the total number of participants are due to missing data points for some participants.

TABLE 2 | Directly assessed difficulties by experimental condition.

Sample	Task type	Condition	Smiley	χ^2	df	p	Navigation	χ^2	df	p
1	Quantity correspondence	Verbal	46/46	1.01	1	0.315	45/46	6.03	1	0.014
		Non-verbal	45/46				38/46			
	Quantity comparison	Verbal	46/46	1.01	1	0.315	46/46	1.01	1	0.315
		Non-verbal	45/46				45/46			
	Ordering	Verbal	46/46	1.03	1	0.309	46/46	1.06	1	0.304
		Non-verbal	44/45				43/44			
2	Non-symbolic addition	Verbal	70/70	3.02	1	0.082	69/70	5.82	1	0.016
		Non-verbal	68/71				62/70			
	Non-symbolic subtraction	Verbal	70/70	n.a.			68/69	1.02	1	0.312
		Non-verbal	71/71				70/70			
	Cross-modal addition	Verbal	70/70	1.02	1	0.312	68/69	1.04	1	0.309
		Non-verbal	71/71				70/70			
	Symbolic arithmetic	Verbal	69/69	n.a.			68/68	n.a.		
		Non-verbal	71/71				71/71			

n.a., not applicable due to 1-level factor.

Summarily, we observed that all but a few participants had correctly understood the feedback symbols and the navigation arrow from the start.

Indirectly Assessed Difficulties: Practice Repetition

As an indirect measure of usability, we examined whether the number of participants that repeated the practice session of each task differed by experimental condition. **Table 3** presents contingency tables and χ^2 -tests of association. **Figure 2** presents percentage of repeaters per condition and task.

TABLE 3 | Indirectly assessed difficulties (practice repetition) by experimental condition.

Sample	Task type	Condition	Repeater		χ^2	df	p
			No	Yes			
1	Quantity correspondence	Verbal	27	19	0.55	1	0.46
		Non-verbal	33	17			
	Quantity comparison	Verbal	38	8	7.90	1	0.005
		Non-verbal	28	22			
	Ordering	Verbal	4	42	51.70	1	<0.001
		Non-verbal	41	9			
2	Non-symbolic addition	Verbal	47	23	5.81	1	0.016
		Non-verbal	60	11			
	Non-symbolic subtraction	Verbal	47	23	3.14	1	0.076
		Non-verbal	57	14			
	Cross-modal addition	Verbal	30	40	6.82	1	0.009
		Non-verbal	46	25			
	Symbolic arithmetic	Verbal	53	17	0.27	1	0.6
		Non-verbal	51	20			

The number of participants that repeated the practice session did not vary significantly between conditions in the *Quantity correspondence task*, the *Non-symbolic subtraction task* and the *Symbolic arithmetic task*. Fewer participants repeated the practice session in the non-verbal condition of the *Ordering*, *Non-symbolic addition* and *Cross-modal addition* tasks. Inversely, more participants repeated the practice session in the non-verbal condition of the quantity comparison task.

Task Descriptives

Internal Consistency

Internal consistency of the experimental tasks in the first sample ranged from good to questionable (see **Table 4**). Only the *Ordering* task with non-symbolic answers showed unacceptable internal consistency. Due to the low number of items in each task, we estimated internal consistency without differentiation as to answer format in the second sample. While the *Symbolic arithmetic* task provided acceptable (Subtraction) to good (Addition) internal consistency, the three other tasks only reached poor to questionable consistency.

Tests for Normality

All task scores showed ceiling effects (somewhat less pronounced in Sample 2), independently from experimental group or the symbolic nature of the task, thus deviating significantly from the normal distribution (statistical tests for all subtests are reported in **Table 4**). Skewed distributions were expected considering the test was designed to differentiate at the bottom end of the ability distribution. Consequently, the Shapiro-Wilks tests showed substantial non-normality. Therefore, we conducted non-parametric analysis of variance to examine possible group differences in task performance.

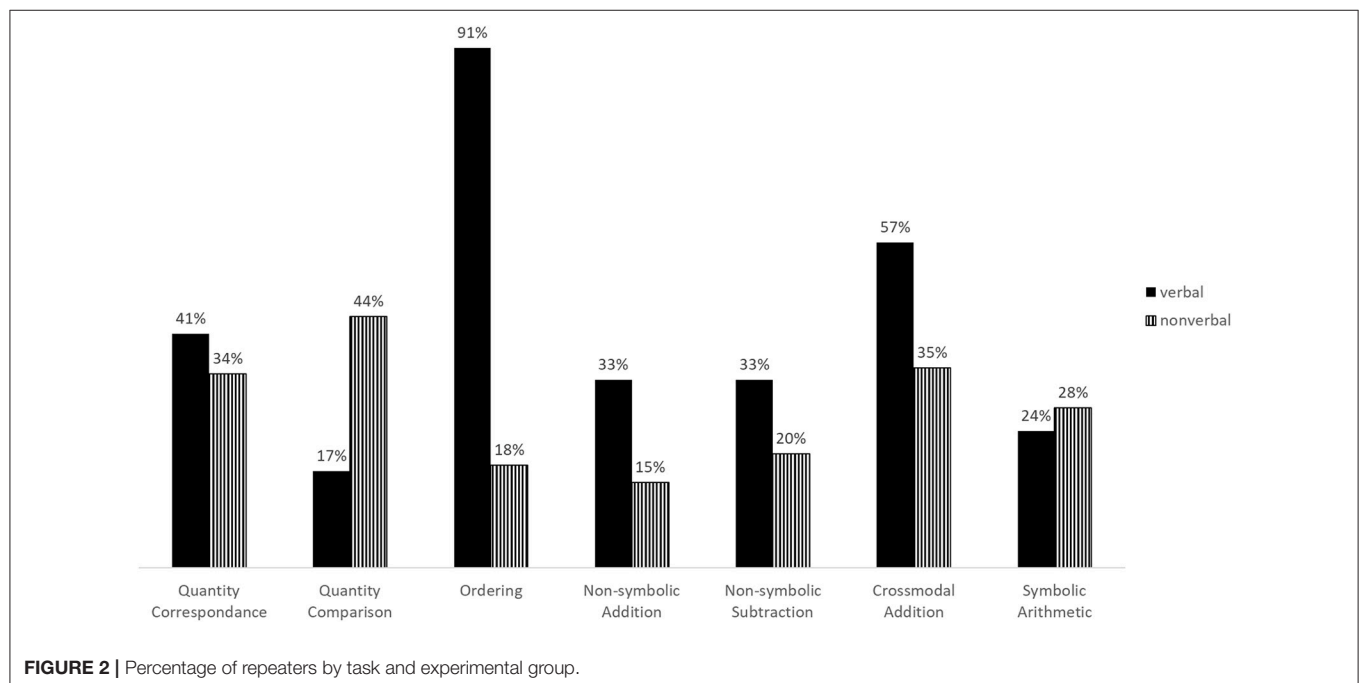


TABLE 4 | Task performance, descriptives and non-verbal vs. verbal comparison.

Task type	Cond.	N	M(POMP) (SD)	Range		Internal consistency		Skewness	S-W		ANOVA (K-W)	
				Theor.	Emp.	α	ω		W	p	χ^2	P
SAMPLE 1												
Quantity correspondence (NS)	Verbal	46	0.86 (0.19)	0–1	0.08–1	0.813	0.852	–1.77	0.69	<0.001	1.47	>0.05
	Non-verbal	50	0.91 (0.16)	0–1	0.17–1			–2.54				
Quantity correspondence (S)	Verbal	46	0.96 (0.09)	0–1	0.63–1	0.746	0.819	–2.39	0.48	<0.001	1.63	>0.05
	Non-verbal	50	0.92 (0.17)	0–1	0.25–1			–2.63				
Quantity comparison (NS)	Verbal	46	0.95 (0.13)	0–1	0.50–1	0.892	0.899	–2.94	0.48	<0.001	1.62	>0.05
	Non-verbal	50	0.87 (0.27)	0–1	0–1			–1.94				
Quantity comparison (S)	Verbal	46	0.96 (0.14)	0–1	0.25–1	0.697	0.737	–3.54	0.48	<0.001	3.70	>0.05
	Non-verbal	50	0.88 (0.24)	0–1	0–1			–2.00				
Ordering (NS)	Verbal	46	0.78 (0.26)	0–1	0.25–1	0.462	0.521	–0.68	0.83	<0.001	0.60	>0.05
	Non-verbal	50	0.74 (0.25)	0–1	0–1			–0.80				
Ordering (S)	Verbal	46	0.92 (0.22)	0–1	0–1	0.735	0.763	–2.78	0.53	<0.001	3.02	>0.05
	Non-verbal	50	0.88 (0.23)	0–1	0–1			–2.08				
SAMPLE 2												
Non-symbolic addition (NS)	Verbal	46	0.78 (0.24)	0–1	0–1	0.495	0.553	–0.82	0.75	<0.001	1.94	>0.05
	Non-verbal	50	0.84 (0.19)	0–1	0.33–1			–0.69				
Non-symbolic addition (S)	Verbal	46	0.66 (0.31)	0–1	0–1			–0.40	0.84	<0.001	0.15	>0.05
	Non-verbal	50	0.67 (0.32)	0–1	0–1			–0.63				
Non-symbolic subtraction (NS)	Verbal	46	0.88 (0.21)	0–1	0–1	0.593	0.618	–1.88	0.61	<0.001	0.00	>0.05
	Non-verbal	50	0.89 (0.19)	0–1	0.33–1			–1.51				
Non-symbolic subtraction (S)	Verbal	46	0.62 (0.35)	0–1	0–1			–0.50	0.83	<0.001	1.27	>0.05
	Non-verbal	50	0.69 (0.33)	0–1	0–1			–0.78				
Cross-modal addition (NS)	Verbal	46	0.79 (0.26)	0–1	0–1	0.439	0.480	–0.90	0.75	<0.001	0.03	>0.05
	Non-verbal	50	0.79 (0.27)	0–1	0–1			–1.07				
Cross-modal addition (S)	Verbal	46	0.62 (0.33)	0–1	0–1			–0.49	0.86	<0.001	1.37	>0.05
	Non-verbal	50	0.57 (0.31)	0–1	0–1			–0.28				
Symbolic arithmetic (Add.)	Verbal	46	0.95 (0.14)	0–1	0–1	0.880	0.888	–4.76	0.32	<0.001	0.07	>0.05
	Non-verbal	50	0.94 (0.21)	0–1	0–1			–4.05				
Symbolic arithmetic (Sub.)	Verbal	46	0.85 (0.23)	0–1	0–1	0.787	0.803	–2.04	0.66	<0.001	0.01	>0.05
	Non-verbal	50	0.83 (0.27)	0–1	0–1			–1.84				

POMP, Percentage of maximum performance; S-W, Shapiro-Wilk test of normality; K-W, Kruskal-Wallis ANOVA on ranks.

Scale Intercorrelations

In Sample 1 performances on almost all experimental tasks correlated significantly among each other (see **Table 5**). The exception was the *Quantity comparison* task (symbolic format), which did not correlate significantly with the *Quantity correspondence* task (non-symbolic format) and with the *Ordering* task (both formats).

The reported correlations in the following paragraph are all significant (see **Table 6**). Letters in parentheses indicate the answer format (NS = non-symbolic; (S) = symbolic). In Sample 2, performances in *Symbolic arithmetic* (addition and subtraction) correlated with each other and with performance in all other tasks having a symbolic response format (i.e. *Non-symbolic addition*, *Non-symbolic subtraction*, and *Cross-modal addition*). Performance in *Symbolic arithmetic* did not correlate with performance in tasks requiring non-symbolic output, except for the *Non-symbolic subtraction* task. Performances in *Non-symbolic addition* and *subtraction* (S) correlated with performance on all other tasks. Performances in the two *Non-symbolic arithmetic* (NS) did not correlate with each

other. Performance in *Cross-modal addition* (S) correlated with performance in all other tasks, except *Non-symbolic arithmetic* (i.e., *Non-symbolic addition* and *Non-symbolic subtraction*) with non-symbolic response formats. Performance in *Cross-modal addition* (NS) correlated with performance in all other tasks, except *Symbolic arithmetic*.

Criterion Validity

In Sample 1, average performance (all experimental tasks combined) correlated significantly with all criterion validity tasks (see **Table 7**) except with the two-digit SYMP test.

In Sample 2, average performance (all experimental tasks combined) correlated significantly with all criterion validity tasks.

Comparison of Task Performance: Verbal vs. Non-verbal Instructions

Analyses of variance (Kruskal-Wallis) on task scores with experimental group (verbal vs. non-verbal) as between-subjects factor revealed no significant differences in any of the

TABLE 5 | Scale intercorrelations: Sample 1.

Scale intercorrelations Sample 1		Quantity correspondence (S)	Quantity comparison (NS)	Quantity comparison (S)	Ordering (NS)	Ordering (S)
Quantity correspondence (NS)	Rho	0.516	0.424	0.189	0.436	0.296
	p	<0.001	<0.001	0.065	<0.001	0.003
Quantity correspondence (S)	Rho		0.374	0.212	0.294	0.215
	p		<0.001	0.038	0.004	0.036
Quantity comparison (NS)	Rho			0.612	0.443	0.381
	p			<0.001	<0.001	<0.001
Quantity comparison (S)	Rho				0.125	0.147
	p				0.225	0.153
Ordering (NS)	Rho					0.519
	p					<0.001

S, Symbolic answer format; NS, Non-symbolic answer format; Rho, Spearman's rho.

TABLE 6 | Scale intercorrelations: Sample 2.

Scale intercorrelations Sample 2		NS Add. (S)	NS Sub. (NS)	NS Sub. (S)	Cross. Add. (NS)	Cross. Add. (S)	Sym. Arith. (Add)	Sym. Arith. (Sub)
Non-symbolic addition (NS)	Rho	0.257	0.157	0.317	0.306	0.162	0.138	0.046
	p	0.002	0.063	<0.001	<0.001	0.056	0.102	0.590
Non-symbolic addition (S)	Rho		0.335	0.372	0.244	0.260	0.236	0.417
	p		<0.001	<0.001	0.003	0.002	0.005	<0.001
Non-symbolic subtraction (NS)	Rho			0.342	0.197	0.042	0.241	0.231
	p			<0.001	0.019	0.619	0.004	0.006
Non-symbolic subtraction (S)	Rho				0.249	0.290	0.193	0.352
	p				0.003	<0.001	0.022	<0.001
Cross-modal addition (NS)	Rho					0.301	0.091	0.165
	p					<0.001	0.282	0.051
Cross-modal addition (S)	Rho						0.207	0.211
	p						0.014	0.012
Symbolic arithmetic (addition & subtraction) (S)	Rho							0.262
	p							0.002

(S), Symbolic answer format; (NS), Non-symbolic answer format; Rho, Spearman's rho.

TABLE 7 | Criterion validity.

Criterion validity		TTR+	TTR–	Counting (oral)	Counting (written)	SYMP (one digit)	SYMP (two digit)
SAMPLE 1							
Average test score (all tasks)	Rho	0.453	0.349	0.279	0.475	0.308	0.111
	p	<0.001	<0.001	0.006	<0.001	0.002	0.28
SAMPLE 2							
Average test score (all tasks)	Rho	0.431	0.355	0.441	0.499	0.409	0.26
	p	<0.001	<0.001	<0.001	<0.001	<0.001	0.002

Rho, Spearman's rho.

tasks, neither in Sample 1 nor in Sample 2 (see Table 4). Overall performances were very high in the non-verbal and in the verbal condition (ranging between 57 and 96%), indicating that children succeeded comparably well in both conditions.

DISCUSSION

The purpose of the present study was to explore the possibility of measuring basic math competence in young children without using verbal instructions. To this aim we developed a series

of computerized tasks presented on tablet-computers either verbally, using traditional language instructions or non-verbally, using video instructions repeatedly showing successful task completion and assessed whether the instruction type influenced task performance.

Usability Aspects

To check whether this new mode of instruction was effective, we assessed the comprehensibility of the tasks both directly and indirectly. Regarding the prior, the feedback symbols (the green happy and the red sad smiley faces during the instruction and practice phase) were easily understood by most if not all participants. The same is true for the navigation symbol (the arrow to both save the answer and switch to the next item).

As an indirect assessment of task comprehension, we examined differences in the number of participants that repeated the practice session of each task. Given the low difficulty level of the tasks presented during instruction and practice, we assumed that children who did not get the practice items right in their first attempt had not understood the purpose of the task at first and therefore needed a second run. In three tasks [*Quantity correspondence* (S1), *Non-symbolic subtraction* (S2), and *Symbolic arithmetic* (S2)], the number of repeaters did not vary significantly, suggesting that non-verbal instructions can be understood as well as verbal ones. On the other hand, we observed significantly less repeaters in three other tasks [*Non-symbolic addition* (S2), *Ordering* (S1) and *Cross-modal addition* (S2)] when children were instructed non-verbally, implying that non-verbal instructions can be more effective than verbal ones in these situations. This tendency was especially pronounced in the *Ordering* task. Finally, we found an inverse difference in repeaters in the *Quantity comparison* task. Significantly more participants repeated the practice session of the *Quantity comparison* task when they received non-verbal instructions. Conveying “choose the side that has more” through a video showing successful task completion repeatedly seems to have worked less well than simply giving the participants an explicit verbal instruction to do so, even though we displayed more repetitions in this task than in the other tasks. This shows that not every task instruction can be easily replaced by non-verbal videos without adding unnecessary complexity. This result stands in stark contrast with our observations concerning the *Ordering* task, which was understood much better following non-verbal instructions. Because the verbal instruction requested to order items from left to right, the extreme difference in repeaters (91% vs. 18%) could possibly be attributed to the fact that reliable left /right distinction has not been achieved by children of this age. Notwithstanding, this observation illustrates well that a single word in the instruction can lead to a complete failure to understand the task at hand and that this can be easily avoided by using non-verbal video instructions. Taken together, our results based on the repetition of practice items suggest that non-verbal instructions are an efficient alternative to the classically used verbal instructions and might in some cases even be more direct and effective. However, they do not provide a universally applicable solution, because on rare occasions they fail to convey task instructions as clearly and unequivocally as spoken language.

Anecdotally, it appeared that children were generally highly motivated to complete our tasks and many asked if they could do them again. This might be due to the video-game-like appearance of the assessment tool, which differs considerably from the paper-and-pencil material that they encounter in everyday math classes, which probably helped to promote task compliance and motivation (Lumsden et al., 2016).

Validity Aspects

Scale intercorrelations indicate that performance in the three tasks assessed in Sample 1 (i.e., *Quantity correspondence*, *Quantity comparison*, *Ordering*) largely correlated, which may reflect the fact that they rely, at least in part, on the same basic numerical competences. While performance on the non-symbolic version of the *Quantity comparison* task did correlate with performance on most other experimental tasks, performance on the symbolic version of the *Quantity comparison* task shows less consistent correlations with performance on other tasks. Most strikingly, the latter does not correlate significantly with performance on the *Ordering* task, both symbolic and non-symbolic versions. This stands in contrast with most findings in recent literature that report strong correlation between performance on tasks measuring cardinality (*Quantity comparison* task) and ordinality (*Ordering* task) (e.g., Lyons et al., 2014; Sasanguie et al., 2017; Sasanguie and Vos, 2018). This might be due to reporting correlations for the whole sample without distinguishing instruction type: a large proportion of participants in the video condition of the task did not seem to correctly understand its purpose, which could explain the absence of correlation between its performance and any other task. Accordingly, the *Quantity comparison* task will need to be adapted in future studies. Sample 2 consisted of calculation tasks that were either presented in classical symbolic or more unusual non-symbolic and/or cross-modal format (i.e., *Symbolic addition and subtraction*, *Non-symbolic addition and subtraction*, *Cross-modal addition*). In this sample, performance in symbolic arithmetic correlated with performance in those tasks having a symbolic response format, but not those requiring non-symbolic answers. This points toward a special role of number symbol processing, in line with the importance of this ability for mathematics (e.g., Bugden and Ansari, 2011; Bugden et al., 2012). Interestingly, and in line with the importance of number symbols, performance in non-symbolic arithmetic tasks with symbolic output formats also correlated with all calculation tasks of Sample 2. While validating the main expectations concerning our task and their properties, conclusions concerning scale intercorrelations remain provisional at this stage, since all tasks could not be correlated with each other in the present design due to two different participant samples.

Considering the overall medium reliability of our experimental tasks, special care should be taken to include more items assessing performance in the different tasks in further developments of this project.

Finally, we observed that average performance of all experimental tasks combined correlated significantly with performance in most (Sample 1) to all (Sample 2) control tasks. The control tasks were chosen to cover the most established

measures of basic math competences in young children, known to predict latter differences in standardized mathematical tests and distinguish children with MLD from typically developing peers. We therefore included tasks assessing children's abilities to count (Goldman et al., 1988; Geary et al., 1999; Passolunghi and Siegel, 2004; Willburger et al., 2008; Hornung et al., 2014), to compare symbolic magnitudes (De Smedt et al., 2009, 2013; Brankaer et al., 2017) and to calculate (De Vos, 1992; Geary et al., 1993; Klein and Bisanz, 2000; Locuniak and Jordan, 2008; Geary, 2010). The non-significant correlation between performance of the tasks in the first sample with performance in the two-digit symbolic number comparison task can be attributed to participant's lack of knowledge on two-digit numbers at the time of data collection (approx. 5 weeks of schooling) (MENFP, 2011; Martin et al., 2013).

Task Performance Compared by Experimental Group

Type of instruction prior to the test did not affect participants' performance in any of the experimental tasks. We observed high average performance in both samples and similar performances in both experimental conditions. This leads us to conclude that instruction type does not seem have an observable effect on future task performance. In other words, explicit verbal instructions can be replaced by videos showing successful task completion for children to understand the functioning and purpose of the numerical and mathematical tasks. This is an important result when put in the context of multilingual settings in particular, where the language of instruction can have considerable negative effects on task performance. Indeed, video instructions seem to work as well as traditional verbal instructions while taking language out of the equation.

At this point, we want to stress that we do not claim that mathematics and language can be assessed independently (Dowker and Nuerk, 2016). Indeed, prior research has shown that while the logic and procedures of counting are stored independently from language, the learning of even small number words relies on linguistic skills (Wagner et al., 2015). Also, languages inverting the order of units and tens in number words negatively affect the learning of number concepts and arithmetic (Zuber et al., 2009; Göbel et al., 2014; Imbo et al., 2014). Other studies have highlighted that proficiency in the language of instruction (Abedi and Lord, 2001; Hickendorff, 2013; Paetsch et al., 2016; Saalbach et al., 2016) and, more specifically, the mastery of mathematical language are essential predictors of mathematics performance (Purpura and Reid, 2016). It also becomes increasingly clear that test language modulates the neuronal substrate of mathematical cognition (Salillas and Carreiras, 2014; Salillas et al., 2015; Van Rinsveld et al., 2017). On the other hand, we do claim that a testee's access to the assessment tools should not be limited by proficiency in a certain language. Although most existing tasks already use images to minimize linguistic load, they still rely on some form of verbal instruction or vocabulary that needs to be fully understood to solve the task correctly. We thus think that it is not sufficient to minimize language load in mathematics items, but that it

would be preferential to remove linguistic demands altogether. Our results show that this can be achieved by using implicit video instructions that rely on participant's non-verbal cognitive skills.

Limitations and Future Studies

A first limitation for the interpretation of our results are the medium internal consistency scores of many of our tasks. We aimed to explore as many tasks as possible using non-verbal instructions, while keeping total test time under 40 min due to children's limited attention span (Manly et al., 2001). This led to some psychometric compromises by offering only a few items per task and subscale (i.e., symbolic and non-symbolic answer format), especially for the tasks in the second sample. In the future, we will select the tasks with the highest potential of differentiating in the lower spectrum of ability and supplement them with more items.

To further differentiate experimental conditions, it would have been possible to present only word problems and exclude all animations in the verbal instruction group whenever possible. For example, instead of showing pigs moving into a barn, the animation could be replaced with a written/spoken story on pigs going into a barn before offering three possible answers. We expect that such a contrasted design would lead to more significant differences in task comprehension and would be particularly interesting to investigate differences in item functioning in relationship to the participant's language background. In order to provide a robust proof of concept for the valid use of video instructions we decided here to adapt a more conservative approach with minimal differences between the video and verbal conditions. However, it would be interesting to use also more contrasted conditions in future studies.

Additionally, we anecdotally observed that touchscreen responsiveness seemed to be an issue with more impulsive participants. Indeed, when the touchscreen did not react to a first touch by showing a bold border around the selected image, these participants switched to another answer. We speculate that they interpreted the non-response of the tool as a wrong answer on their part and choose to try another one. This is an unfortunate but important technical limitation that will be addressed in future versions of the application, as impulsivity and attention issues are strongly correlated with mathematical abilities, especially in the target population for this test (LeFevre et al., 2013). Finally, we want to stress the difference in participant's age between the two sets of tasks presented here. In future developments of this project, homogenous groups of children from the first half of the first grade should be targeted.

CONCLUSION

Taken together, these preliminary results show that explicit verbal instructions do not seem to be required for assessing basic math competencies when replaced by instructional videos. While variations depending on the task and the quality of experimental instructions are present, video instructions seem to constitute a valid alternative to traditional verbal instructions. In addition, the video-game-like aspect of the present assessment tool was well

received, contributing positively to children's task compliance and motivation. All in all, the results of this study provide an important and encouraging proof of concept for further developments of language neutral and fair tests without verbal instructions.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the research ethics guidelines by the ethics review panel of the University of Luxembourg and has been approved by the former. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

REFERENCES

- Abedi, J., and Lord, C. (2001). The language factor in mathematics education. *Appl. Measur. Educ.* 14, 219–234. doi: 10.1207/S15324818AME1403_2
- Aster, M. G., von, Bzufka, M. W., and Horn, R. R. (2009). ZAREKI-K. *Neuropsychologische Testbatterie für Zahlenverarbeitung und Rechnen bei Kindern: Kindergartenversion: Manual*. Frankfurt: Harcourt.
- Barth, H., Kanwisher, N., and Spelke, E. (2003). The construction of large number representations in adults. *Cognition* 86, 201–221. doi: 10.1016/S0010-0277(02)00178-6
- Braden, J. P., and Hannah, J. M. (1998). "9 - Assessment of hearing-impaired and deaf children with the WISC-III A2 - Prifitera, Aurelio," in *Practical Resources for the Mental Health Professional*, eds D. H. B. T.-W.-I. C. U., and I. Saklofske (San Diego, CA: Academic Press), 175–201.
- Brankaer, C., Ghesquière, P., and De Smedt, B. (2017). Symbolic magnitude processing in elementary school children: a group administered paper-and-pencil measure (SYMP Test). *Behav. Res. Methods* 49, 1361–1373. doi: 10.3758/s13428-016-0792-3
- Bugden, S., and Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition* 118, 35–47. doi: 10.1016/j.cognition.2010.09.005
- Bugden, S., Price, G. R., McLean, D. A., and Ansari, D. (2012). Developmental cognitive neuroscience the role of the left intraparietal sulcus in the relationship between symbolic number processing and children's arithmetic competence. *Acc. Anal. Prevent.* 2, 448–457. doi: 10.1016/j.dcn.2012.04.001
- Butterworth, B., Varma, S., and Laurillard, D. (2011). Dyscalculia: from brain to education. *Science* 332, 1049–1053. doi: 10.1126/science.1201536
- Cattell, R. B., and Cattell, A. K. S. (1973). *Culture Fair Intelligence Tests: CFIT*. Institute for Personality & Ability Testing.
- Cohen, P., Cohen, J., Aiken, L. S., and West, S. G. (1999). The problem of units and the circumstance for POMP. *Multiv. Behav. Res.* 34, 315–346. doi: 10.1207/S15327906MBR3403_2
- Courtney, A. S., Hayes, F. B., Couch, K. W., and Frick, M. (1984). Administration of the WISC-R performance scale to hearing-impaired children using pantomimed instructions. *J. Psychoeduc. Assess.* 2, 1–7. doi: 10.1177/073428298400200101
- De Smedt, B., Noël, M. P., Gilmore, C., Ansari, D., Noël, M.-P., Gilmore, C., et al. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trend. Neurosci. Educ.* 2, 48–55. doi: 10.1016/j.tine.2013.06.001
- De Smedt, B., Verschaffel, L., and Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *J. Exp. Child Psychol.* 103, 469–479. doi: 10.1016/j.jecp.2009.01.010
- De Vos, T. (1992). *Tempo-Test-Rekenen. Handleiding. [Tempo Test Arithmetic. Manual]*. Nijmegen: Berkhout.

AUTHOR CONTRIBUTIONS

MG and CS were responsible for the conception and design of the study. MG was responsible for the acquisition, analysis and interpretation of the data as well as the drafting of the paper. CH, TB, CM, RM, and CS made critical contributions to the interpretation of the data and the revision of the draft.

FUNDING

The research presented in this paper is being funded by Luxembourg's National Research Fund (FNR) under grant N° 10099885.

- Desoete, A., Ceulemans, A., Roeyers, H., and Huylebroeck, A. (2009). Subitizing or counting as possible screening variables for learning disabilities in mathematics education or learning? *Educ. Res. Rev.* 4, 55–66. doi: 10.1016/j.edurev.2008.11.003
- Dowker, A., and Nuerk, H.-C. (2016). Linguistic influences on mathematics. *Front. Psychol.* 7:1035. doi: 10.3389/fpsyg.2016.01035
- Feis, Y. F. (2010). "Raven's progressive matrices," in *Encyclopedia of Cross-Cultural School Psychology*, ed C. S. Clauss-Ehlers (New York, NY: Springer), 787.
- Gallagher-Mitchell, T., Simms, V., and Litchfield, D. (2017). Learning from where 'eye' remotely look or point: impact on number line estimation error in adults. *Q. J. Exp. Psychol.* 71, 1526–1534. doi: 10.1080/17470218.2017.1335335
- Geary, D. C. (2010). Mathematical disabilities: reflections on cognitive, neuropsychological, and genetic components. *Learn. Individ. Differ.* 20, 130–133. doi: 10.1016/j.lindif.2009.10.008
- Geary, D. C., Cormier, P., Goggin, J. P., Estrada, P., and Lunnn, M. C. E. (1993). *Speed-of-Processing Across Monolingual, Weak Bilingual, And Strong Bilingual Adults*.
- Geary, D. C., Hoard, M. K., and Hamson, C. O. (1999). Numerical and arithmetical cognition: patterns of functions and deficits in children at risk for a mathematical disability. *J. Exp. Child Psychol.* 74, 213–239. doi: 10.1006/jecp.1999.2515
- Gersten, R., Jordan, N. C., and Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *J. Learn. Disabil.* 38, 293–304. doi: 10.1177/00222194050380040301
- Göbel, S. M., Moeller, K., Pixner, S., Kaufmann, L., Nuerk, H.-C. C., et al. Nuerk, H.-C. C. (2014). Language affects symbolic arithmetic in children: the case of number word inversion. *J. Exp. Child Psychol.* 119, 17–25. doi: 10.1016/j.jecp.2013.10.001
- Goldman, S. R., Pellegrino, J. W., and Mertz, D. L. (1988). Extended practice of basic addition facts: strategy changes in learning-disabled students. *Cogn. Instr.* 5, 223–265. doi: 10.1207/s1532690xci0503_2
- Haag, N., Heppt, B., Roppelt, A., and Stanat, P. (2014). Linguistic simplification of mathematics items: effects for language minority students in Germany. *Eur. J. Psychol. Educ.* 30, 145–167. doi: 10.1007/s10212-014-0233-6
- Haag, N., Heppt, B., Stanat, P., Kuhl, P., and Pant, H. A. (2013). Second language learners' performance in mathematics: disentangling the effects of academic language features. *Learn. Inst.* 28, 24–34. doi: 10.1016/j.learninstruc.2013.04.001
- Haffner, J., Baro, K., Parzer, P., and Resch, F. (2005). *Heidelberger Rechentest (HRT 1-4) [Heidelberg Calculation Test]*. Göttingen: Hogrefe.
- Hickendorff, M. (2013). The language factor in elementary mathematics assessments: computational skills and applied problem solving in a multidimensional IRT framework. *Appl. Measur. Educ.* 26, 253–278. doi: 10.1080/08957347.2013.824451
- Hornung, C., Schiltz, C., Brunner, M., and Martin, R. (2014). Predicting first-grade mathematics achievement: the contributions of domain-general cognitive abilities, nonverbal number sense, and early number competence. *Front. Psychol.* 5:272. doi: 10.3389/fpsyg.2014.00272

- Imbo, I., Vanden Bulcke, C., De Brauwier, J., and Fias, W. (2014). Sixty-four or four-and-sixty? The influence of language and working memory on children's number transcoding. *Front. Psychol.* 5:313. doi: 10.3389/fpsyg.2014.00313
- Jordan, N. C., Glutting, J., and Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learn. Individ. Differ.* 20, 82–88. doi: 10.1016/j.lindif.2009.07.004
- Klein, J. S., and Bisanz, J. (2000). Preschoolers doing arithmetic: The concepts are willing but the working memory is weak. *Can. J. Exp. Psychol. / Revue Can. de Psychol. Exp.* 54:105.
- LeFevre, J., Fast, L., Skwarchuk, S., Smith-Chant, B. L., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: longitudinal predictors of performance. *Child Dev.* 81, 1753–1767. doi: 10.1111/j.1467-8624.2010.01508.x
- LeFevre, J.-A., Berrigan, L., Vendetti, C., Kamawar, D., Bisanz, J., Skwarchuk, S.-L., et al. (2013). The role of executive attention in the acquisition of mathematical skills for children in Grades 2 through 4. *J. Exp. Child Psychol.* 114, 243–261. doi: 10.1016/j.jecp.2012.10.005
- Lenz, T. (2015). *Bildungsbericht Luxemburg*. Luxembourg: University of Luxembourg.
- Locuniak, M. N., and Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *J. Learn. Disabil.* 41, 451–459. doi: 10.1177/0022219408321126
- Lohman, D. F., and Hagen, E. P. (2001). *Cognitive Abilities Test (Form 6)*. Rolling Meadows, IL: Riverside.
- Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., and Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: a systematic review of applications and efficacy. *JMIR Serious Games* 4:e11. doi: 10.2196/games.5888
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., and Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Dev. Sci.* 17, 714–726. doi: 10.1111/desc.12152
- Manly, T., Anderson, V., Nimmo-Smith, I., Turner, A., Watson, P., and Robertson, I. H. (2001). The differential assessment of children's attention: the Test of Everyday Attention for Children (TEA-Ch), normative sample and ADHD performance. *J. Child Psychol. Psychiatry* 42, 1065–1081. doi: 10.1111/1469-7610.00806
- Martin, R., Ugen, S., and Fischbach, A. (2013). *Épreuves Standardisées – Bildungsmonitoring Luxemburg*. Luxembourg: University of Luxembourg.
- MENFP (2011). *Plan d'études: Ecole fondamentale*.
- Ministère de l'éducation nationale de l'enfance et de la Jeunesse. (2015). *Luxembourgish Education System in Key Figures School year 2014/2015*. Luxembourg.
- Naglieri, J. A. (2003). "Naglieri Nonverbal Ability Tests," in *Handbook of Nonverbal Assessment*, ed R. S. McCallum (Boston, MA: Springer), 175.
- Noël, M.-P., Grégoire, J., and Nieuwenhoven, V. (2008). *Test Diagnostique des Compétences de Base en Mathématiques*. Paris: Editions du Centre de Psychologie appliquée (E.C.P.A.).
- Nosworthy, N., Bugden, S., Archibald, L., Evans, B., and Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLoS ONE* 8:e67918. doi: 10.1371/journal.pone.0067918
- Ortiz, S. O., and Dynda, A. M. (2005). "Use of intelligence tests with culturally and linguistically diverse populations," in *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, eds D. P. Flanagan and P. L. Harrison (New York, NY: Guilford Press), 545–556.
- Paetsch, J., Radmann, S., Felbrich, A., Lehmann, R., and Stanat, P. (2016). Sprachkompetenz als Prädiktor mathematischer Kompetenzentwicklung von Kindern deutscher und nicht-deutscher Familiensprache. *Z. Entwicklungspsychol. Pädagog. Psychol.* 48, 27–41. doi: 10.1026/0049-8637/a000142
- Passolunghi, M. C., and Siegel, L. S. (2004). Working Memory and access to numerical information in children with disability in mathematics. *J. Exp. Child Psychol.* 88, 348–367. doi: 10.1016/j.jecp.2004.04.002
- Pellegrini, A. D., and Bohn, C. M. (2005). The role of recess in children's cognitive performance and school adjustment. *Educ. Res.* 34, 13–19. doi: 10.3102/0013189X034001013
- Purpura, D. J., and Reid, E. E. (2016). Mathematics and language: individual and group differences in mathematical language skills in young children. *Early Child. Res. Q.* 36, 259–268. doi: 10.1016/j.ecresq.2015.12.020
- Ricken, G., Fritz, A., and Balzer, L. (2011). Mathematik und Rechnen – Test zur Erfassung von Konzepten im Vorschulalter (MARKO-D) – ein Beispiel für einen niveaorientierten Ansatz. *Empirische Sonderpädagogik* 3, 256–271.
- Saalbach, H., Gunzenhauser, C., Kempert, S., and Karbach, J. (2016). Der Einfluss von Mehrsprachigkeit auf mathematische Fähigkeiten bei Grundschulkindern mit niedrigem sozioökonomischen Status. *Frühe Bildung* 5, 73–81. doi: 10.1026/2191-9186/a000255
- Salillas, E., and Carreiras, M. (2014). Core number representations are shaped by language. *Cortex* 52, 1–11. doi: 10.1016/j.cortex.2013.12.009
- Salillas, E., Barraza, P., and Carreiras, M. (2015). Oscillatory brain activity reveals linguistic prints in the quantity code. *PLoS ONE* 10:e0121434. doi: 10.1371/journal.pone.0121434
- Sasanguie, D., and Vos, H. (2018). About why there is a shift from cardinal to ordinal processing in the association with arithmetic between first and second grade. *Dev. Sci.* 21:e12653. doi: 10.1111/desc.12653
- Sasanguie, D., Lyons, I. M., De Smedt, B., and Reynvoet, B. (2017). Unpacking symbolic number comparison and its relation with arithmetic in adults. *Cognition* 165, 26–38. doi: 10.1016/j.cognition.2017.04.007
- Sasanguie, D., Van Den Bussche, E., and Reynvoet, B. (2012). Predictors for mathematics achievement? Evidence from a longitudinal study. *Minds Brain Educ.* 6, 119–128. doi: 10.1111/j.1751-228X.2012.01147.x
- Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science* 174, 1285–1295. doi: 10.1126/science.174.4016.1285
- Schaupp, H., Holzer, N., and Lenart, F. (2007). ERT 1+. Eggenberger Rechentest 1+. *Diagnostikum Für Dyskalkulie Für Das Ende Der*, 1.
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20, 1–16. doi: 10.1111/desc.12372
- van Luit, J. E. H., van de Rijt, B. A. M., and Hasemann, K. (2001). *Osnabrücker Test zur Zahlbegriffsentwicklung: OTZ*. Hogrefe, Verlag für Psychologie.
- Van Rinsveld, A., Dricot, L., Guillaume, M., Rossion, B., and Schiltz, C. (2017). Mental arithmetic in the bilingual brain: language matters. *Neuropsychologia* 101, 17–29. doi: 10.1016/j.neuropsychologia.2017.05.009
- Van Rinsveld, A., Schiltz, C., Brunner, M., Landerl, K., Ugen, S., Rinsveld, A., et al. (2016). Solving arithmetic problems in first and second language: does the language context matter? *Learn. Inst.* 42, 72–82. doi: 10.1016/j.learninstruc.2016.01.003
- Wagner, K., Kimura, K., Cheung, P., and Barner, D. (2015). Why is number word learning hard? Evidence from bilingual learners. *Cogn. Psychol.* 83, 1–21. doi: 10.1016/j.cogpsych.2015.08.006
- Willburger, E., Fussenegger, B., Moll, K., Wood, G., and Landerl, K. (2008). Naming speed in dyslexia and dyscalculia. *Learn. Individ. Differ.* 18, 224–236. doi: 10.1016/j.lindif.2008.01.003
- Zuber, J., Pixner, S., Moeller, K., and Nuerk, H. C. (2009). On the language specificity of basic number processing: transcoding in a language with inversion and its relation to working memory capacity. *J. Exp. Child Psychol.* 102, 60–77. doi: 10.1016/j.jecp.2008.04.003

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

Copyright © 2018 Greisen, Hornung, Baudson, Muller, Martin and Schiltz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Individuality in the Early Number Skill Components Underlying Basic Arithmetic Skills

Jonna B. Salminen^{1*}, Tuire K. Koponen² and Asko J. Tolvanen³

¹ Department of Education, Special Education, University of Jyväskylä, Jyväskylä, Finland, ² Faculty of Education and Psychology, Centre for Research on Learning and Teaching, University of Jyväskylä, Jyväskylä, Finland, ³ Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland

OPEN ACCESS

Edited by:

Annemie Desoete,
Ghent University, Belgium

Reviewed by:

Bert Reynvoet,
KU Leuven, Belgium
Lars Orbach,
Universität Duisburg-Essen, Germany

*Correspondence:

Jonna B. Salminen
jonna.b.salminen@jyu.fi

Specialty section:

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

Received: 23 February 2018

Accepted: 05 June 2018

Published: 02 July 2018

Citation:

Salminen JB, Koponen TK and
Tolvanen AJ (2018) Individuality
in the Early Number Skill Components
Underlying Basic Arithmetic Skills.
Front. Psychol. 9:1056.
doi: 10.3389/fpsyg.2018.01056

Early number skills underlie success in basic arithmetic. However, very little is known about the skill profiles among children in preprimary education and how the potential profiles are related to arithmetic development. This longitudinal study of 440 Finnish children in preprimary education (mean age: 75 months) modeled latent performance-level profile groups for the early number skill components that are proposed to be key predictors of arithmetic (symbolic number comparison, mapping, and verbal counting skills). Based on three assessment time points (September, January, and May), four profile groups were found: the poorest-performing (6%), low-performing (16%), near-average-performing (33%), and high-average-performing children (45%). Although the differences between the groups were statistically significant in all three number skill components and in basic arithmetic, the poorest-performing children seemed to have serious difficulties in accessing the semantic meaning of symbolic numbers that was required in the number comparison and mapping tasks in this study. Interestingly, the tasks demanding processing between quantities and symbols also most differentiated the poorest-performing children from the low-performing children. Due to remarkable and stable individual differences in early number skill components, the findings suggest systematic support and progress monitoring practices in preeducational settings to diminish and avoid potential difficulties in arithmetic and mathematics in general.

Keywords: early number skill components, arithmetic, preprimary education, latent profile analysis, poorest-performing children, low-performing children

INTRODUCTION

Typically, as an innate ability, children are able to quickly discriminate small sets of quantities without counting (1–4; subitizing range), and they can detect which of two presented quantities is larger if the difference between them is large enough (Dehaene, 2011; see also von Aster, 2000; von Aster and Shalev, 2007). It has been proposed that this ability is critical for the development of early number skills and especially for number concept skills for which children need to learn the quantitative meaning of small number words (one, two, and three; Butterworth, 2005), and later on, to map verbal and quantitative representations to corresponding number symbols. Along these skills, children recite number words very early (Fuson, 1988; Wynn, 1990; Krajewski and Schneider, 2009) which forms a base for learning exact verbal counting list (Fuson, 2009) and for enumerating and calculating quantities above the aforementioned subitizing range. To enumerate quantities correctly, children need to master and follow the procedural principles for counting

(Gelman and Gallistel, 1978; one-to-one correspondence, stable order of the counting words, and cardinality). Children also need to understand what can be counted and that the order in which the quantities are counted does not matter (Gelman and Gallistel, 1978). These principles are vital for exact object counting (see also Krajewski and Schneider, 2009; Dehaene, 2011), which, in turn, relates to the development of number concept skills. Thus, understanding the association between different numerical representations that are number words, quantities, and Arabic number symbols plays a critical role in the development of early number skills (Krajewski and Schneider, 2009; Geary, 2013). Furthermore, this skill allows and strengthens the understanding of explicit number system (knowing the exact relationships between numbers) that can be seen as prerequisite for the ability to compose and decompose magnitudes and for learning efficient and flexible arithmetical calculation strategies (Krajewski and Schneider, 2009; Geary, 2013).

Atypicalities in number skills development and lack of early numerical experiences, as well as math language, increase the risk of facing challenges in learning arithmetic and mathematics at school. One main feature in mathematical learning difficulties (MD) is dysfluency in calculation skills that is deficit in arithmetic fact retrieval (Geary, 2011). That is why researchers try to draw a theoretical picture of number skills development and specify the critical early components related to arithmetic. It has been proposed that the strongest predictor of fluent arithmetic may be symbolic number processing skills (Bartelet et al., 2014; Skagerlund and Träff, 2014; De Smedt, 2015; Vanbinst et al., 2015). Children with MD might have deficits in accessing the numerical meaning from Arabic number symbols (assessed typically by number comparison task; *which of the two number symbols is larger*) which could then be related to basic arithmetic skills and math achievement in general (Rousselle and Noël, 2007; De Smedt and Gilmore, 2011; De Smedt, 2015). On the other hand, deficit in symbolic number processing might become visible in mapping task where fluent ability to transcode between non-symbolic and symbolic numerical notations is required. Previous research has shown that symbolic number comparison and mapping are separable although correlated skills, and mapping is related to mathematics achievement over and above numerical magnitude comparison skills (Brankaer et al., 2014). Deficits in mapping could also explain difficulties in understanding number relations (Geary, 2013). Finally, it has also been proposed that verbal counting plays an important role as a predictor of fluent arithmetic (e.g., Aunola et al., 2004; Zhang et al., 2014; Koponen et al., 2016), and could be a core component in identifying children with potential MD.

Before formal schooling, children typically use counting based strategies for solving simple sums and ease their counting by using manipulatives, fingers, and/or verbal counting. Later, counting strategies develop (through counting all – counting on – counting on from larger number) in consequence of repetitions and routines which in turn allow children to strengthen associations between arithmetical problems and their solutions (Peters and De Smedt, 2018). That is why verbal counting might play an important and foundational role in learning arithmetic.

As known, individual differences in early number skills appear to be relatively stable and the differences widen in subsequent years (e.g., Aunola et al., 2004; Desoete and Grégoire, 2006; Murphy et al., 2007; Geary et al., 2008; Morgan et al., 2009, 2011; Wong et al., 2014). To better understand the potential qualitative differences between the poorest-performing and low-performing children (Geary, 2011), and to give targeted support for individual needs (Dowker and Sigley, 2010) we need specific knowledge of children's skill-profiles in separate number skill components. To date, mostly two types of studies have examined these early number skill components: studies on a certain factor (Price and Wilkey, 2017; Vanbinst et al., 2018) and studies on composite scores (Jordan et al., 2006, 2007; Aunio et al., 2015). The first type tries, more or less, to deepen the knowledge of the core factors of MD but typically does not simultaneously model two or more core components at the same time. The second type tries, more or less, to understand the developmental trajectories of number skills underlying fluent arithmetic. Neither approach allows us to draw a clear picture of how these early number skill components are related and what kind of skill profiles may exist at kindergarten age before formal schooling.

To conclude, defining a clear picture of the underlying components predicting arithmetic skills is challenging due to the varying approaches, measures, sampling issues, and age levels used in previous studies (see De Smedt et al., 2013; Lyons et al., 2014; Hart et al., 2016). Thus, we need specific knowledge of the individuality in the number skill components. This question is not only theoretically interesting but also provides new information for planning and suggesting reasonable, targeted support to prevent persistent deficits and cumulative difficulties in mathematics (Butterworth et al., 2011; Geary, 2011). One way to examine the individual differences in theoretically distinct and unique contributors of basic arithmetic is to use person-oriented analysis methods (Bartelet et al., 2014; Skagerlund and Träff, 2014). This approach tries to get support and add potential new knowledge for existing theories by driving the data instead of differentiating groups of children who are clustered with certain cut-off thresholds. The present study implemented latent profile analysis method (LPA) to investigate the heterogeneity of potential early skill profiles in the three number skill components strongly underlying fluent arithmetic: symbolic number processing (NC), mapping (MS), and verbal counting skills (VC). Along these skills, non-symbolic magnitude comparison and number line acuity also predict arithmetic achievement. However, symbolic number comparison correlates more strongly with math achievement than non-symbolic number comparison (for review see Schneider et al., 2017; within kindergarteners see Sasanguie et al., 2012). In addition, instead seeing number line acuity as a direct predictor of math skills it should be seen as a factor influencing on the developmental process of both skills (e.g., Friso-van den Bos et al., 2015). The main interest of the current study was to get more evidence of the early number skill components that challenge especially the poorest- and low-performing children the most and that are measurable for practitioners in small-group conditions by paper-and-pencil tasks.

This longitudinal study aimed first to examine whether different performance-level profile groups in early number skill components are found among children in preprimary education (research question 1, RQ1). The second aim was to examine which of the components potentially differentiates the profile groups the most (RQ2). The third aim was to examine whether the preprimary education group, gender, or age plays a role in belonging to a certain profile group (RQ3). Finally, the between-group differences in basic arithmetic were tested (RQ4). The three screening tools with negatively skewed distribution were used to assess early number skill components in September, January, and May. With this procedure, the study aimed to deepen knowledge of the skill performance of poorly performing children through the preprimary education year (for researchers) and to reliably screen children in need of extra support for numerical skills (for practitioners). Therefore, differentiation of performance levels among typical-, average-, or high-achieving children was not the focus. The theoretical model and the three main research questions are presented later in **Figure 1**.

information letter with the descriptions of our study purpose, procedure, and contact information. Parents were informed of their right to decline or discontinue the children's participation to our study at any time point. Parents were also informed that we will not ask any information to identify children from the data (such as surname, birth date, etc.) and therefore, written permissions from parents to us were not required. The final sample sizes varied from 486 to 557 kindergarteners, depending on the assessment point and given the option for teachers, parents, and children to commence or cease participation at any point. Altogether, 30 teachers and 440 kindergarteners who participated in all three assessment time points were included in the analyses. These longitudinal data for Finnish kindergarteners were geographically representative, and when tested, participant attrition was not found to be systematically related to any of the early number skill components assessed in this study. The final sample consisted of 223 girls (*mean age* = 75.19 months, *SD* = 3.58) and 215 boys (*mean age* = 74.94 months, *SD* = 3.75) and two other children with missing gender information.

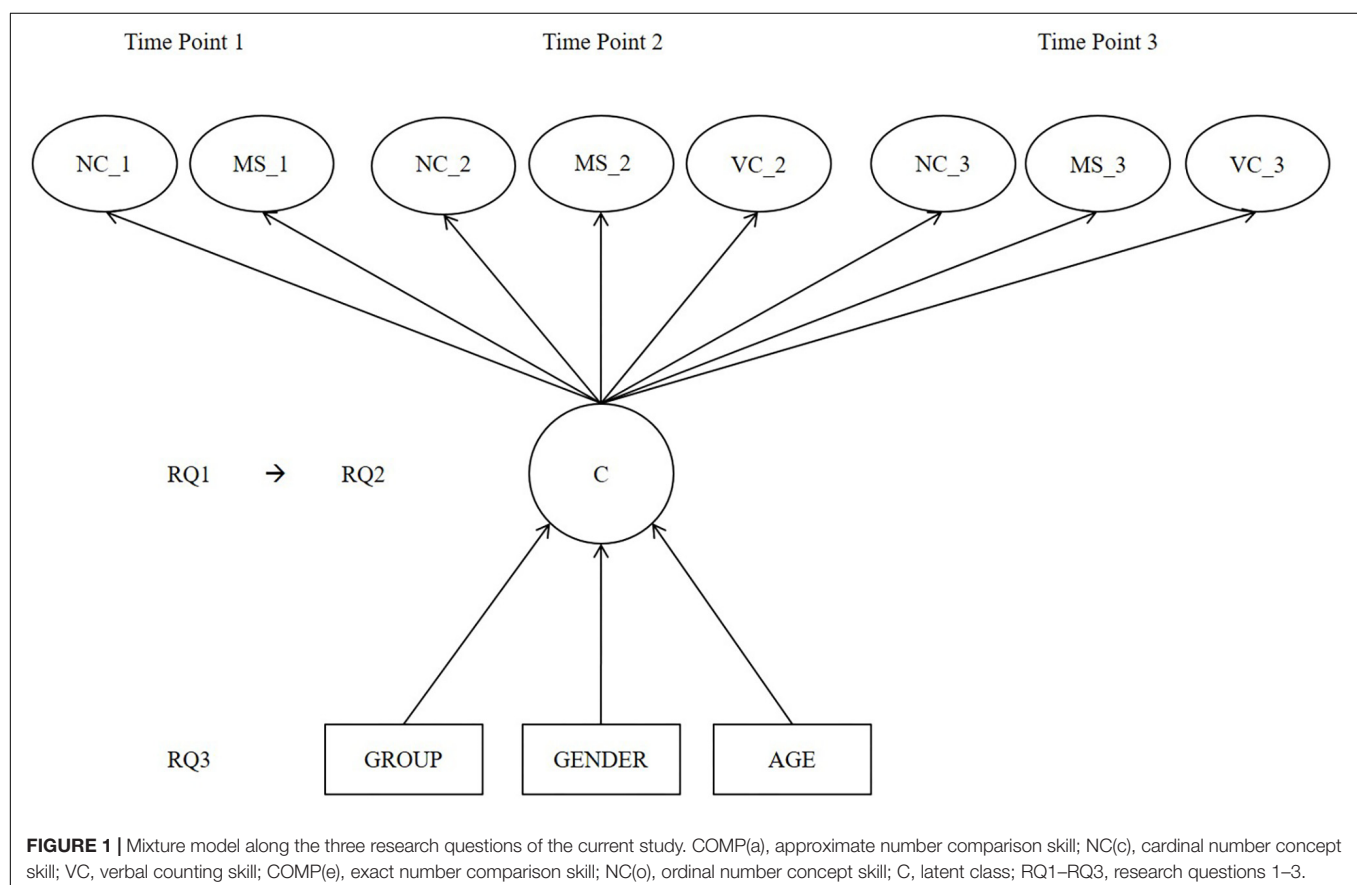
MATERIALS AND METHODS

Participants

At the outset, 35 kindergarten teachers voluntarily participated in the study as data collection coordinators. Parents received an

Procedure

In August, the volunteer teachers were trained for the three-tiered group assessment procedure, which was conducted in September, January, and May during the preprimary education year. The following number skill components were assessed:



symbolic number comparison (September, January, and May), mapping (September, January, and May), and verbal counting skills (January and May). The tools were piloted before the actual study to ensure that the participating teachers were able to follow the instructions for the assessment procedure, and that the items would measure expected dimensions and cover individuality of skill-levels. The teachers then administered the assessment procedure within their own preprimary education group during small group sessions (of 5–8 children). The teachers instructed the tasks item by item to the children who responded by cross marking one of the three alternatives presented on the paper. After each assessment point, the teachers returned all materials for each assessment point to the research assistants who were trained to work with the data. This procedure was carried out at each assessment time point (September, January, and May). After each assessment point, we tested the validity and difficulty of items. Based on the results and expected skills development, the amount of items were reduced and changed and new skill components added to the following screener for receiving meaningful variance.

With individual and small group assessment settings (attention), and with permission to repeat the instructions (working memory), as well as by using multiple-choice, paper-and-pencil items without time limits (response inhibition), the demand for executive functioning skills during the assessments was thought to be diminished. By varying and challenging the number skill components over time (i.e., changing numerical distances among alternatives, growing the number area, and adding assessed components), the difficulty level was thought to increase from fall to winter to spring. With this decision, practitioners could screen weaknesses at different cross-sectional time points by comparing the individual performance levels to typically developing children with diminished risk of a potential floor or ceiling effect or a test–retest effect.

Measures

Symbolic Number Comparison

Symbolic number comparison skill was assessed at time point 1 (eight items, Cronbach's $\alpha = 0.88$) and time point 2 (six items, Cronbach's $\alpha = 0.91$) in small group settings. At both time points, the first half of the assessment tasks included items from which the child was asked to choose and mark the largest written number among three alternatives, presented horizontally (e.g., 9, 4, and 7). The second half consisted of tasks in which the child was asked to choose the smallest written number (e.g., 6, 10, and 8). Each item was coded as zero (incorrect) or 1 point (correct) or as an empty cell (missing value), so that the approximate number comparison formed a categorical variable for the analysis. At time point 3, the number comparison task required exact comparison skill (four items, Cronbach's $\alpha = 0.59$). The child was asked to choose which of the three alternatives included one more, two more, one fewer, and two fewer than the item originally presented. Each item was coded as zero (incorrect) or 1 point (correct) or as an empty cell (missing value) and was set as categorical items for the analysis to first evaluate their validity and difficulty level in assessing number comparison skills. Based

on the item difficulty analysis (see section “Data Analysis” and the Appendix), three NC variables, one per time point, were included in the final analysis as parceled variables (NC_1, NC_2, and NC_3).

Mapping

Mapping skills were assessed at time point 1 (16 items, Cronbach's $\alpha = 0.88$) and time point 2 (eight items, Cronbach's $\alpha = 0.71$) in small group settings. The test included four types of tasks each consisting of four items (time point 1) or two items (time point 2). For each task type, the child was asked to choose the corresponding numerical representation from among three alternatives. First, number words were contrasted with quantities (dots), and then number words were contrasted with written number symbols (e.g., the number word “eight” was said aloud and the written symbols 7, 9, and 8 were presented), then, quantities were contrasted with written symbols (without verbal hints), and finally, written symbols were contrasted with quantities. Each item was coded as zero (incorrect) or 1 point (correct) or as an empty cell (missing value) so that cardinal number concept skill formed a categorical variable for the analysis. At time point 3, the task consisted only of four items (Cronbach's $\alpha = 0.60$) in which the child was asked to mark the 12th, the 17th, every 2nd, and finally, every 3rd item among several alternatives presented horizontally for each task. Each item was coded as zero (incorrect) or 1 point (correct) or as an empty cell (missing value) and was set as categorical items for the analysis to first evaluate their validity and difficulty level in assessing mapping skills. Based on item difficulty analysis (see section “Data Analysis” and the Appendix), three MS variables, one per time point, were included in the final analysis as parceled variables (MS_1, MS_2, and MS_3).

Verbal Counting

Verbal counting was assessed individually at time point 2 and at time point 3 with identical tasks (nine items, Cronbach's $\alpha = 0.84$ and 0.82 , respectively). First, the child was asked to count forward starting from one. This task was divided into three subtasks: to count correctly up to the number word 10, to the number word 20, and to the number word 30. Second, the child was asked to count backward, again in three subtasks: to count backward correctly from 5 to 1, from 12 to 8, and from 20 to 16. Third, the child was asked to skip count by twos, again in three subtasks: to count correctly up to the number word 10, to the number word 18, and to the number word 30. Each item was coded as zero (incorrect) or 1 point (correct) or as an empty cell (missing value) and was set as categorical items for the analysis to first evaluate their validity and difficulty level in assessing verbal counting skills. Based on item difficulty analysis (see section “Data Analysis” and the Appendix), two VC variables, one per time point, were included in the final analysis as parceled variables (VC_2 and VC_3; numbers indicating the time point).

Basic Arithmetic Story Problems

Basic arithmetic was assessed at time point 3 (eight items, Cronbach's $\alpha = 0.63$) in small group settings. Four of the

assessment tasks were verbally presented addition tasks (*A boy has three fishes. He gets two more fishes. How many fishes does he have now?*), in which the children needed to give their responses by marking the correct number symbol among three alternatives presented horizontally. With a similar procedure, the child was asked to respond to four other tasks that were subtraction tasks (*A girl has five keys. She gives two keys away. How many keys does she have now?*). Each item was coded as zero (incorrect) or 1 point (correct) or as an empty cell (missing value). Because this study focused on the prerequisite skills for arithmetic (NC, MS, and VC), this task was included in *post hoc* analysis only as a sum score of eight items for testing potential differences in basic arithmetic between hypothetically meaningful profile groups.

Data Analysis

First, item response theory (IRT) analysis was needed and conducted for each of the eight number skill components: to assess the items' ability to measure the dimensions and cover all the individuals' skills level, to evaluate item difficulties, and factor loadings. Model parameters were estimated using the weighted least squares means and variance adjusted (WMLSV estimator) estimation in Mplus version 7.11 (Muthén and Muthén, 1998–2012). Goodness-of-fit was evaluated based on the following criteria: chi-square test of model fit (χ^2), root-mean-square error of approximation (RMSEA), comparative fit index (CFI), Tucker-Lewis index (TLI), and weighted root-mean-square residual (WRMR). Values for well-fitting measurement models were as follows: RMSEA < 0.06, CFI > 0.95, TLI > 0.95, and WRMR < 0.09. To reduce the number of estimated parameters for the sample size, parcels were formed using item difficulty information from the IRT analysis. The classification of individual items into the parcels was also based on content (e.g., different types of verbal counting items, including counting on, counting backward from a given number, and counting on by twos, were mixed in each verbal counting parcel to add balance among the three parcels). The goodness-of-fit with the estimates RMSEA, TLI, CFI, and WRMR for different types of latent number skill components are presented in the Appendix along the item difficulty information and factor loadings per dimension (NC_1, MS_1, and VC_2). The time points, when the components were assessed the first time, were used because the following components were formed from the originally presented items (i.e., the following assessment points contained an equal

or smaller number of items compared to previous assessment points). To better evaluate the validity of the NC and MS components at time point 3 (because the reported Cronbach alpha values were relatively small probably due to the small number of items, 0.59 and 0.60, respectively), the factor loadings for these dimensions (NC_3 and MC_3) are also presented separately in the Appendix. Correlations between the eight latent number skill factors are presented in **Table 1** for the whole sample ($N = 440$). Based on the measurement models, factor scores were computed for use in the second step of the analysis. Item difficulty, standardized factor loadings of each item, and the parceling information are presented in the Appendix.

Second, LPA across a total of eight latent number component factor scores was used to empirically identify potential skill profile groups (RQ1). Mplus provides several statistical fit indices for deciding the number of latent classes. In the present study, individuals ($N = 440$) were classified into different latent profile groups using the following criteria: the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the adjusted BIC, the entropy index, average posterior probabilities, and statistical test results for the Lo-Mendell-Rubin Likelihood ratio test (LMRL), Lo-Mendell-Rubin test (LMR), and bootstrap likelihood ratio test (BLRT). As the three screening tools were developed to differentiate the potential skill levels of poorly performing children and their potential differences on separate number skill components (RQ2), LPA was terminated when the average posterior probabilities and class counts proposed new groups of near-average- and/or high-average-performing children with small class counts. Analyses for between-profile-group differences in terms of preprimary education group, age, and gender (RQ3) were conducted using the auxiliary option in Mplus (Muthén and Muthén, 1998–2012). Finally, for testing potential group differences in BA, the independent samples *t*-test was used (RQ4).

RESULTS

Research Question 1

In LPA, the parsimonious number of classes was four with class counts of 25 (the poorest-performing; 6%), 71 (low-performing; 16%), 147 (near-average-performing; 33%), and 197 (high-average-performing; 45%) when all eight latent basic number skill components were included in the analysis (**Figure 2**).

TABLE 1 | Correlations between early number skill component factor scores.

Factor	1	2	3	4	5	6	7
(1) Number Comparison (NC_1)							
(2) Mapping Skills (MS_1)	0.695						
(3) Number Comparison (NS_2)	0.723	0.637					
(4) Mapping Skills (MS_2)	0.533	0.785	0.604				
(5) Verbal Counting (VC_2)	0.556	0.710	0.706	0.616			
(6) Number Comparison (NC_3)	0.616	0.713	0.715	0.677	0.725		
(7) Mapping Skills (MS_3)	0.535	0.684	0.764	0.619	0.639	0.764	
(8) Verbal Counting (VC_3)	0.516	0.710	0.665	0.612	0.907	0.733	0.684

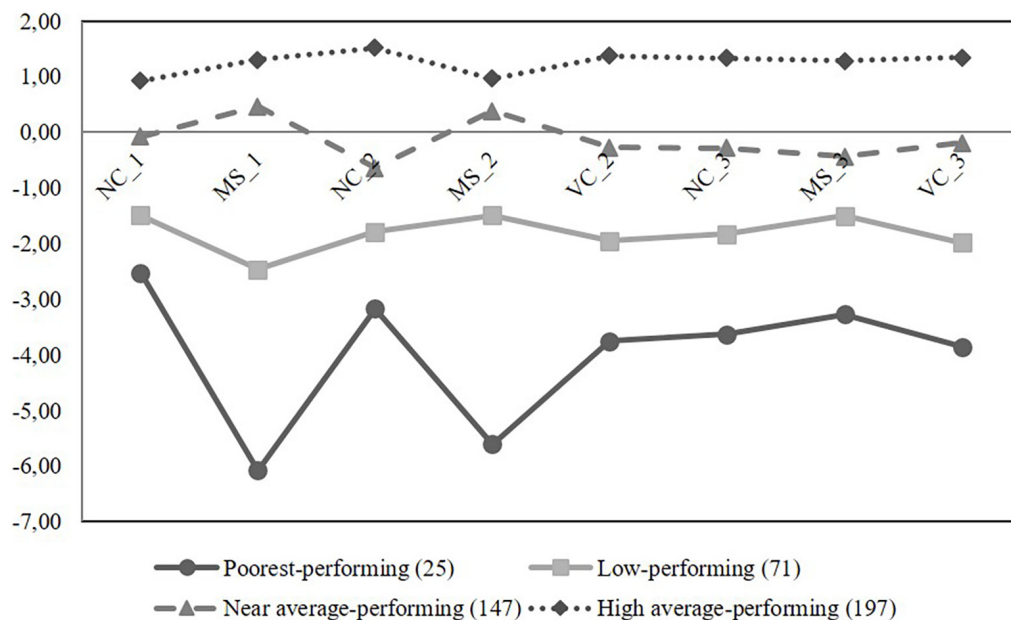


FIGURE 2 | Latent profile groups across three time points with standardized estimates for intercept of each latent factor. COMP(a), approximate number comparison skill; NC(c), cardinal number concept skill; VC, verbal counting skill; COMP(e), exact number comparison skill; NC(o), ordinal number concept skill; 1–3, time points 1–3.

Average latent class probabilities for most likely latent class membership were 0.999 for the poorest-performing group, 0.964 for the low-performing group, 0.954 for the near-average-performing group, and 0.970 for the high-average-performing group indicating very high stability of group membership. Model fit indices for different class solutions are presented in **Table 2**.

Research Question 2

Based on confidence interval comparisons, all four profile groups differed statistically significantly from each other on every latent skill component over the preprimary education year (**Table 3**). Further, the poorest-performing children performed equally poorly in number comparison and mapping tasks while for the other groups of children mapping task seemed to somewhat be easier than number comparison task. The percentages of accuracy were 35% for the poorest-, 51% for the low-, 79% for the near-average-, and 97% for the high-average-performing children in number comparison task. The respective percentages

were 38, 69, 93, and 98% for the group of poorest-, low-, near-average-, and high-average-performing children in mapping. That is why mapping skill seemed to most differentiate the poorest-performing children from the other profile groups (**Figure 2**). In more detail, the items that required mapping between quantities and written number symbols and vice versa were the most difficult for the poorest-performing children. The percentages of correctly mapped numerical representations in the poorest-performing group were as follows: 23% for quantities to number symbols and vice versa and 54% for number words to quantities and vice versa. The respective percentages were 56 and 81% for the group of low-performing children; 90 and 96% for the near-average-; and 97 and 99% for the high-average-performing children.

Research Question 3

There were no between-group differences in terms of participating in preeducation instruction in a certain

TABLE 2 | Standardized fit indices for latent profile analysis over early number skill components.

No. of classes	LL	No. of free parameters	AIC	BIC	Adj. BIC	Entropy	VLMR	Adj. VLMR	BLRT
2	−4040.83	25	8131.65	8233.82	8154.49	0.96	$p < 0.001$	$p < 0.001$	$p < 0.001$
3	−3661.77	34	7391.55	7530.50	7422.60	0.93	$p = 0.054$	$p = 0.056$	$p < 0.001$
4	−3429.83	43	6945.66	7121.39	6984.93	0.94	$p < 0.001$	$p < 0.001$	$p < 0.001$
5	−3313.62	52	6731.25	6943.76	6778.74	0.94	$p = 0.085$	$p = 0.089$	$p < 0.001$
6	−3238.58	61	6599.16	6848.45	6654.87	0.95	$p = 0.225$	$p = 0.230$	$p < 0.001$

LL, log-likelihood; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; Adj., adjusted; VLMR, Vuong-Lo-Mendell-Rubin test; BLRT, Bootstrap Likelihood-Ratio-Test. The best-fitting solution is shown in boldface.

TABLE 3 | Standardized estimates for intercepts with confidence intervals in four-class solution over early number skill components.

Latent variable	Latent group											
	Poorest-performing (25)			Low-performing (71)			Near average-performing (147)			High average-performing (197)		
	<i>i</i> (SE)	<i>CI</i> (99%)		<i>i</i> (SE)	<i>CI</i> (99%)		<i>i</i> (SE)	<i>CI</i> (99%)		<i>i</i> (SE)	<i>CI</i> (99%)	
		Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper
Time point 1												
Number comparison	−2.53 (0.27)	−2.15	−1.351	−1.49 (0.20)	−1.349	−0.71	−0.08 (0.12)	−0.28	0.17	0.92 (0.06)	0.57	0.69
Mapping skills	−6.08 (0.54)	−3.17	−2.33	−2.46 (0.26)	−1.40	−0.83	0.46 (0.11)	0.09	0.32	1.30 (0.10)	0.55	0.63
Time point 2												
Number comparison	−3.17 (0.29)	−2.12	−1.40	−1.79 (0.16)	−1.21	−0.78	−0.64 (0.12)	−0.55	−0.17	1.52 (0.10)	0.76	0.93
Mapping skills	−5.61 (0.56)	−3.56	−2.36	−1.49 (0.25)	−1.10	−0.47	0.38 (0.08)	0.11	0.30	0.96 (0.08)	0.45	0.56
Verbal counting	−3.75 (0.32)	−2.38	−1.77	−1.95 (0.21)	−1.32	−0.84	−0.27 (0.14)	−0.34	0.05	1.37 (0.11)	0.67	0.84
Time point 3												
Number comparison	−3.63 (0.33)	−2.44	−1.69	−1.83 (0.18)	−1.27	−0.81	−0.29 (0.10)	−0.32	−0.01	1.33 (0.12)	0.64	0.87
Mapping skills	−3.27 (0.32)	−2.40	−1.55	−1.50 (0.13)	−1.09	−0.71	−0.44 (0.10)	−0.41	−0.11	1.28 (0.11)	0.64	0.90
Verbal counting	−3.86 (0.34)	−2.47	−1.79	−1.98 (0.20)	−1.34	−0.85	−0.19 (0.13)	−0.30	0.09	1.34 (0.09)	0.67	0.81

i, intercept; CIs, confidence intervals.

kindergarten group ($n = 30$). However, according to the chi-square test with basic precursors, the high-average-performing children were statistically significantly older (mean age, 75.84 months, $SE = 0.25$) than the children in the three other groups (the poorest-performing $\text{mean} = 73.87$ months, $SE = 0.78$, chi-square = 5.90, $p = 0.015$; low-performing $\text{mean} = 74.04$ months, $SE = 0.48$, chi-square = 11.28, $p = 0.001$; near-average-performing $\text{mean} = 74.69$ months, $SE = 0.31$, chi-square = 7.97, $p = 0.005$). Finally, there appeared to be more boys within the poorest-performing group than in the near- (chi-square = 6.85, $p = 0.009$) or high-average-performing (chi-square = 7.18, $p = 0.007$) groups but not compared to the low-performing group. To conclude, the poorest- and low-performing profile groups did not differ in terms of kindergarten group, age, or gender.

Research Question 4

Latent profile analysis method showed that the poorest- and low-performing profile groups were unique. To confirm the result, an independent-samples t -test was used to examine the potential group difference in basic arithmetic. According to the t -test, the poorest-performing children performed statistically significantly poorer in basic arithmetic than the low-performing children (the poorest-performing $\text{mean} = 4.61$, $SD = 1.67$; low-performing $\text{mean} = 5.89$, $SD = 1.70$; $t(92) = -3.14$, $p = 0.002$, $d = 0.45$).

DISCUSSION

In the present study, latent profile analysis was used to identify potential performance-level groups among 440 Finnish children (6- to 7-year-olds) with distinct number skill profiles. The performance levels in three number skill components, with which fluent arithmetic skills have typically been predicted, were assessed three times during the preprimary education

semester in September, January, and May. The components were number comparison, mapping between different numerical representations (quantities, number words, and number symbols), and verbal counting.

The results of the present study revealed four types of performance profile groups across number comparison, mapping, and verbal counting skills. There was a statistically significant difference in all number skill components between the poorest- (6%), low- (16%), near-average- (33%), and high-average-performing children (45%). Based on these results, the poorest- and low-performing children seem to need acute support for all early number skill components. In particular, the poorest-performing children seem to need specific training for number comparison and mapping skills. Especially, the task types that required exact mapping of quantities with number symbols, as well as number symbols with quantities were the most difficult for the poorest-performing children. Instead, the percentages of accuracy in tasks dealing with number words (number word–quantity and number word–number symbol mapping) were higher. Moreover, the poorest-performing children differed statistically significantly from low-performing children in basic addition and subtraction story problem-solving skills ($d = 0.45$). The poor performance in the early number and story problem-solving skills indicate a clear risk for arithmetical difficulties especially among the poorest-performing children.

Aligned with previous literature, the LPA in the present study suggests that 96 children (22% of the total sample) performed less well than the near- or high-average-performing children. Of these 96 children, 71 formed one unique profile group (low-performing children, representing approximately 16% of the total sample), and 25 formed another unique profile group (the poorest-performing children, representing approximately 6% of the total sample) with high stability of group membership. These proportions (16 and 6%) seem somewhat to be in line with previous findings that children struggling with math skills could

have different types of growth rates if their initial performance level varies between the 11th and 25th percentiles or falls below the 10th percentile (Murphy et al., 2007; Morgan et al., 2009, 2011; Salaschek et al., 2014). This study offers support for this phenomenon by showing that these performance-level differences already exist before formal schooling. The findings are also in line with previous literature (concerning performance levels) although the present study used only very basic number skill components instead of school mathematics (Murphy et al., 2007; Morgan et al., 2009, 2011; Salaschek et al., 2014). Finally, interestingly, the proportion of the poorest-performing children (6%) found in this study was comparable to the estimated prevalence of children with MD who are typically diagnosed as having deficits in arithmetic fluency at older age levels (varying between at 3–7, 5–7, and 5–8%; Landerl and Moll, 2010; Butterworth et al., 2011; Geary, 2011).

The findings also suggest that the poorest-performing children have serious deficits in all early number skills. Further, the percentages of correctness were at the same level within the poorest-performing children in number comparison (35) and in mapping (38). However, the other groups seemed to perform better in mapping than in number comparison task. The percentages of correctness were 69, 93, and 98 within the low-, near-average-, and high-average-performing children, respectively. In number comparison, the corresponding percentages within the low-, near-average-, and high-average-performing children were 51, 79, and 97, respectively. That is why the mapping task differentiated the poorest-performing children from the other groups the most.

In more detail, in mapping task, the poorest-performing children seemed to have more serious deficits than low-performing children especially in matching written number symbols to the corresponding quantities and vice versa. The poorest-performing children showed less serious deficits when verbal number words were included in the mapping tasks. It follows that these findings cannot be explained (at least not fully) by weak dot counting skills or by verbal deficits, as a comparable performance in that case would have been found in written symbol–quantity and verbal number word–quantity mapping tasks. Moreover, the number word–written symbol mapping task was easier for the poorest-performing children than the written symbol–quantity task. This finding lends further support to the suggestion that the most serious deficits are in finding associations between written number symbols and quantities and thus, support the theoretical hypothesis of children with MD having deficits in accessing numerical meaning from written number symbols (De Smedt and Gilmore, 2011). This was supported also by the fact that tasks dealing with number words were easier for the poorest-performing children. That is why number sense (or module) deficit was not supported in our study.

From the developmental perspective, these findings are in line with previous studies suggesting stable and even increasing differences between the unique poorest- and low-performing profile trends (Murphy et al., 2007; Morgan et al., 2009, 2011; Geary, 2011; Wong et al., 2014). Additionally, the mapping tasks operated with number words are developmentally more familiar

to children at first than the tasks requiring understanding of the direct quantity–symbol relationship without verbal support (Dehaene, 1992; von Aster and Shalev, 2007; Geary, 2013). To link these findings to longitudinal studies focusing approximately on the same age level, this study showed that children's age is positively associated with performance level as was shown in Jordan et al.'s (2006) longitudinal study. Older children may have more experience with numbers and (numerical) language than their younger age peers. Therefore, the differences in readiness to benefit from early instructions and participate in peer discussions can be greater between the age levels at the beginning of formal schooling. In contrast to previous findings (for a review, Jordan et al., 2006, 2007; Devine et al., 2013), boys were overrepresented among the low-performing children in the present study in comparison to near- and high-average-performing children, but the poorest- and low-performing groups did not differ by age or gender. The contradictory findings concerning gender differences in mathematics might be due to the methods used for testing differences (Devine et al., 2013). In general, in population-based studies, there are no clear gender differences in the mean level (for a review, see Hyde et al., 2008; Lindberg et al., 2010), but a difference can be found among lower- or higher-performing children (Devine et al., 2013; Stoet and Geary, 2013).

Implications for Educational Practice

The present findings suggest that theoretically valid screening tools have potential to identify children in need of extra support in early number skill components. Moreover, by assessing number comparison, mapping, and verbal counting, it is possible to identify a subgroup of children, with a corresponding prevalence rate of MD, whose poor number skill performance seems to be stable during the whole preprimary education year. The findings suggest that educational practices for early identification of MD risk and early number skills intervention should focus on the most basic skills, especially on quantity-number symbol mapping skill (and vice versa) which most differentiates the poorest-performing children from low-performing children. The stability of poor performance levels found throughout preprimary education indicates a need for systematic progress monitoring of number skill development, as well as planning and offering appropriate mathematical support at the very beginning of formal schooling or perhaps earlier.

Limitations

Deficits in working memory, language, and visuospatial skills (Raghubar et al., 2010; Geary, 2011), processing speed (Willcutt et al., 2013), and certain domains of executive functioning (Friso-van den Bos et al., 2013; Price and Fuchs, 2016; Price and Wilkey, 2017) are also found to be associated with arithmetic skills or math performance more generally. Thus, in future studies, by controlling for general domain skills (see also Kaufmann et al., 2013) and task-specific requirements (De Smedt et al., 2013; Price and Wilkey, 2017), we could better understand the potential qualitative differences between the poorest-performing and low-performing children (Geary, 2011). We also could better identify those children (most) at risk for MD and likewise, plan meaningfully targeted support for individual needs

(Landerl et al., 2009; Rubinsten and Henik, 2009; Butterworth et al., 2011). Unfortunately, in our study administered by teachers, we could not measure these general cognitive skills. We only tried to minimize the demand of executive functioning skills by using a certain type of assessment procedure (e.g., small group sessions, permission to repeat the instructions, only a cross-marking requirement in responses, and non-speeded tasks).

The study tools were developed and tested for practical use. The aim was to develop a set of screeners that could first identify (alert) children in need of extra evaluation and immediate early number skill support (at the beginning of pre-primary education) and then evaluate the progress (in winter and spring times). For this reason, a larger amount of basic skills' items were included into the first screener and then the amount of these items were reduced for being able to add theoretically and developmentally meaningful skills' items into the following screeners (winter and spring) without increasing the assessment effort. This causes three clear limitations for this study. First, reducing the number of items and by changing the assessed skill components we were not able to analyze the number skill development comprehensively (LPA was used instead of growth curve models). Second, by reducing the number of items, some of the sub-skill dimensions showed low reliability values although the reliability for the three screeners as a whole were relatively high (Cronbach alpha values being 0.91; 0.88; and 0.84 respectively). That is why IRT- and factor analysis were conducted for showing the validity of skill components. Third, we were not able to measure all important skills related to arithmetic development. For instance, number line estimation task (as one of the critical measures) would require careful interpretations of the correctness and would therefore be difficult to conduct in screeners meant for practical use. Further, to assess non-symbolic comparison skills with a paper-and-pencil task (which would have been the case in our study), well-controlled items would have been needed (controlling for instance for area, ratio, distance, and response time). However, as they are important, both skills could be individually assessed for example after a classroom-based screening situation for confirming the skill-levels.

One main criticism of using LPA is that the proposed number of classes may not refer to existing subpopulations within the population (Bauer and Curran, 2004). However, in this study,

the best-fitting solution (four profile groups) and the alternative solutions (five or six profile groups) proposed one clear group of the poorest-performing children, in which the latent early number skill components differ most from the other skill-level groups. Thus, findings concerning the poorest-performing children seemed reliable.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Finnish Advisory Board on Research Integrity. The protocol was approved by the Niilo Mäki Institute. All subjects gave written informed consent in accordance with the Finnish Advisory Board on Research Integrity.

AUTHOR CONTRIBUTIONS

JS main work of the whole paper. TK contribution to context. AT contribution to methods.

FUNDING

The data collection of the study administered during the LukiMat-project (started in 2007) funded by the Ministry of Education and Culture in Finland, and administered in Niilo Mäki Institute. This study has been carried out in the Centre for Research on Learning and Teaching and TK has been financed by the Academy of Finland (No. 292 466 for 2015–2019).

ACKNOWLEDGMENTS

We wish to thank the participating preprimary education children, and their teachers, as well as the parents.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01056/full#supplementary-material>

REFERENCES

- Aunio, P., Heiskari, P., van Luit, J. E. H., and Vuorio, J.-M. (2015). The development of early numeracy skills in kindergarten in low-, average- and high-performance groups. *J. Early Child. Res.* 13, 3–16. doi: 10.1177/1476718X14538722
- Aunola, K., Leskinen, E., Lerkkanen, M.-K., and Nurmi, J.-E. (2004). Developmental dynamics of math performance from preschool to grade 2. *J. Educ. Psychol.* 96, 699–713. doi: 10.1037/0022-0663.96.4.699
- Bartelet, D., Ansari, D., Vaessen, A., and Blomert, L. (2014). Cognitive subtypes of mathematics learning difficulties in primary education. *Res. Dev. Disabil.* 35, 657–670. doi: 10.1016/j.ridd.2013.12.010
- Bauer, D. J., and Curran, P. J. (2004). The integration of continuous and discrete latent variable models: potential problems and promising opportunities. *Psychol. Methods* 9, 3–29. doi: 10.1037/1082-989X.9.1.3
- Brankaer, C., Ghesquière, P., and De Smedt, B. (2014). Children's mapping between non-symbolic and symbolic numerical magnitudes and its association with timed and untimed tests of mathematics achievement. *PLoS One* 9:e93565. doi: 10.1371/journal.pone.0093565
- Butterworth, B. (2005). The development of arithmetical abilities. *J. Child Psychol. Psychiatry* 46, 3–18. doi: 10.1111/j.1469-7610.2005.00374.x
- Butterworth, B., Varma, S., and Laurillard, D. (2011). Dyscalculia: from brain to education. *Science* 332, 1049–1053. doi: 10.1126/science.1201536

- De Smedt, B. (2015). "Individual differences in arithmetic fact retrieval," in *Development of Mathematical Cognition, Neural Substrates and genetic Influences*, eds D. Berch, D. Geary, and K. Mann Koepke (San Diego, CA: Academic Press), 219–243.
- De Smedt, B., and Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *J. Exp. Child Psychol.* 108, 278–292. doi: 10.1016/j.jecp.2010.09.003
- De Smedt, B., Noël, M.-P., Gilmore, C., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends Neurosci. Educ.* 2, 48–55. doi: 10.1016/j.tine.2013.06.001
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N
- Dehaene, S. (2011). *The Number Sense: How The Mind Creates Mathematics*. Oxford: Oxford University Press.
- Desoete, A., and Grégoire, J. (2006). Numerical competence in young children and in children with mathematics learning disabilities. *Learn. Individ. Differ.* 16, 351–367. doi: 10.1016/j.lindif.2006.12.006
- Devine, A., Soltész, F., Nobes, A., Goswami, U., and Szűcs, D. (2013). Gender differences in developmental dyscalculia depend on diagnostic criteria. *Learn. Instr.* 27, 31–39. doi: 10.1016/j.learninstruc.2013.02.004
- Dowker, A., and Sigley, G. (2010). Targeted interventions for children with arithmetical difficulties. *Br. Psychol. Soc. Monogr. Ser.* 2, 65–81. doi: 10.1348/97818543370009X12583699332492
- Friso-van den Bos, I., Kroesbergen, E. H., van Luit, J. E. H., Xenidou-Dervou, I., Jonkman, L. M., van der Schoot, M., et al. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *J. Exp. Child Psychol.* 134, 12–29. doi: 10.1016/j.jecp.2015.02.002
- Friso-van den Bos, I., van der Ven, S. H. G., Kroesbergen, E. H., and van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educ. Res. Rev.* 10, 29–44. doi: 10.1016/j.edurev.2013.05.003
- Fuson, K. C. (1988). *Children's Counting and Concepts on Number*. New York, NY: Springer-Verlag. doi: 10.1007/978-1-4612-3754-9
- Fuson, K. C. (2009). Avoiding misinterpretations of Piaget and Vygotsky: mathematical teaching without learning, learning without teaching, or helpful learning-path teaching? *Cogn. Dev.* 24, 343–361. doi: 10.1016/j.cogdev.2009.09.009
- Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *J. Dev. Behav. Pediatr.* 32, 250–263. doi: 10.1097/DBP.0b013e318209edef
- Geary, D. C. (2013). Early foundations for mathematics learning and their relations to learning disabilities. *Curr. Dir. Psychol. Sci.* 22, 23–27. doi: 10.1177/0963721412469398
- Geary, D. C., Hoard, M. K., Nugent, L., and Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Dev. Neuropsychol.* 33, 277–299. doi: 10.1080/87565640801982361
- Gelman, R., and Gallistel, C. R. (1978). *The Child's Understanding of Number*. Cambridge, MA: Harvard University Press.
- Hart, S. A., Logan, J. A. R., Thompson, L., Kovas, Y., McLoughlin, G., and Petrill, S. A. (2016). A latent profile analysis of math achievement, numerosity, and math anxiety in twins. *J. Educ. Psychol.* 108, 181–193. doi: 10.1037/edu0000045
- Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., and Williams, C. C. (2008). Gender similarities characterize math performance. *Science* 321, 494–495. doi: 10.1126/science.1160364
- Jordan, N. C., Kaplan, D., Locuniak, M. N., and Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learn. Disabil. Res. Pract.* 22, 36–46. doi: 10.1111/j.1540-5826.2007.00229.x
- Jordan, N. C., Kaplan, D., Nabors Oláh, L., and Locuniak, M. N. (2006). Number sense growth in kindergarten: a longitudinal investigation of children at risk for mathematics difficulties. *Child Dev.* 77, 153–175. doi: 10.1111/j.1467-8624.2006.00862.x
- Kaufmann, L., Mazzocco, M. M., Dowker, A., von Aster, M., Göbel, S. M., Grabner, R. H., et al. (2013). Dyscalculia from a developmental and differential perspective. *Front. Psychol.* 4:516. doi: 10.3389/fpsyg.2013.00516
- Koponen, T., Salmi, P., Torppa, M., Eklund, K., Aro, T., Aro, M., et al. (2016). Counting and rapid naming predict the fluency of arithmetic and reading skills. *Contemp. Educ. Psychol.* 4, 83–94. doi: 10.1016/j.cedpsych.2016.02.004
- Krajewski, K., and Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: findings from a 3-year longitudinal study. *J. Exp. Child Psychol.* 103, 516–531. doi: 10.1016/j.jecp.2009.03.009
- Landerl, K., Fussenegger, B., Moll, K., and Willburger, E. (2009). Dyslexia and dyscalculia: two learning disorders with different cognitive profiles. *J. Exp. Child Psychol.* 103, 309–324. doi: 10.1016/j.jecp.2009.03.006
- Landerl, K., and Moll, K. (2010). Comorbidity of learning disorders: prevalence and familial transmission. *J. Child Psychol. Psychiatry* 51, 287–294. doi: 10.1111/j.1469-7610.2009.02164.x
- Lindberg, S. M., Hyde, J. S., Petersen, J. L., and Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychol. Bull.* 136, 1123–1135. doi: 10.1037/a0021276
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., and Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Dev. Sci.* 17, 714–726. doi: 10.1111/desc.12152
- Morgan, P. L., Farkas, G., and Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *J. Learn. Disabil.* 42, 306–321. doi: 10.1177/0022219408331037
- Morgan, P. L., Farkas, G., and Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: who falls increasingly behind? *J. Learn. Disabil.* 44, 472–488. doi: 10.1177/0022219411414010
- Murphy, M. M., Mazzocco, M. M. M., Hanich, L. B., and Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *J. Learn. Disabil.* 40, 458–478. doi: 10.1177/00222194070400050901
- Muthén, L. K., and Muthén, B. O. (1998–2012). *Mplus User's Guide*, 7th Edn. Los Angeles, CA: Muthén & Muthén.
- Peters, L., and De Smedt, B. (2018). Arithmetic in the developing brain: a review of brain imaging studies. *Dev. Cogn. Neurosci.* 30, 265–279. doi: 10.1016/j.dcn.2017.05.002
- Price, G. R., and Fuchs, L. S. (2016). The mediating relation between symbolic and nonsymbolic foundations of math competence. *PLoS One* 11:e0148981. doi: 10.1371/journal.pone.0148981
- Price, G. R., and Wilkey, E. D. (2017). Cognitive mechanisms underlying the relation between nonsymbolic and symbolic magnitude processing and their relation to math. *Cogn. Dev.* 44, 139–149. doi: 10.1016/j.cogdev.2017.09.003
- Raghubar, K. P., Barnes, M. A., and Hecht, S. A. (2010). Working memory and mathematics: a review of developmental, individual difference, and cognitive approaches. *Learn. Individ. Differ.* 20, 110–122. doi: 10.1016/j.lindif.2009.10.005
- Rousselle, L., and Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: a comparison of symbolic vs non-symbolic number magnitude processing. *Cognition* 102, 361–395. doi: 10.1016/j.cognition.2006.01.005
- Rubinsten, O., and Henik, A. (2009). Developmental dyscalculia: heterogeneity may not mean different mechanisms. *Trends Cogn. Sci.* 13, 92–99. doi: 10.1016/j.tics.2008.11.002
- Salaschek, M., Zeuch, N., and Souvignier, E. (2014). Mathematics growth trajectories in first grade: cumulative vs. compensatory patterns and the role of number sense. *Learn. Individ. Differ.* 35, 103–112. doi: 10.1016/j.lindif.2014.06.009
- Sasanguie, D., De Smedt, B., Defever, E., and Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *Br. J. Dev. Psychol.* 30, 344–357. doi: 10.1111/j.2044-835X.2011.02048.x
- Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20, 1–16. doi: 10.1111/desc.12372
- Skagerlund, K., and Träff, U. (2014). Number processing and heterogeneity of developmental dyscalculia: subtypes with different cognitive profiles

- and deficits. *J. Learn. Disabil.* 49, 36–50. doi: 10.1177/0022219414522707
- Stoet, G., and Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: within- and across-nation assessment of 10 years of PISA data. *PLoS One* 8:e57988. doi: 10.1371/journal.pone.0057988
- Vanbinst, K., Ceulemans, E., Ghesquière, P., and De Smedt, B. (2015). Profiles of children's arithmetic fact development: a model-based clustering approach. *J. Exp. Child Psychol.* 133, 29–46. doi: 10.1016/j.jecp.2015.01.003
- Vanbinst, K., Ceulemans, E., Peters, L., Ghesquière, P., and De Smedt, B. (2018). Developmental trajectories of children's symbolic numerical magnitude processing skills and associated cognitive competencies. *J. Exp. Child Psychol.* 166, 232–250. doi: 10.1016/j.jecp.2017.08.008
- von Aster, M. (2000). Developmental cognitive neuropsychology of number processing and calculation: varieties of developmental dyscalculia. *Eur. Child Adolesc. Psychiatry* 9, 41–57. doi: 10.1007/s007870070008
- von Aster, M. G., and Shalev, R. S. (2007). Number development and developmental dyscalculia. *Dev. Med. Child Neurol.* 49, 868–873. doi: 10.1111/j.1469-8749.2007.00868.x
- Willcutt, E. G., Petrill, S. A., Wu, S., Boada, R., DeFries, J. C., Olson, R. K., et al. (2013). Comorbidity between reading disability and math disability: concurrent psychopathology, functional impairment, and neuropsychological functioning. *J. Learn. Disabil.* 46, 500–516. doi: 10.1177/0022219413477476
- Wong, T. T.-Y., Ho, C. S.-H., and Tang, J. (2014). Identification of children with mathematics learning disabilities (MLDs) using latent class growth analysis. *Res. Dev. Disabil.* 35, 2906–2920. doi: 10.1016/j.ridd.2014.07.015
- Wynn, K. (1990). Children's understanding of counting. *Cognition* 36, 155–193. doi: 10.1016/0010-0277(90)90003-3
- Zhang, X., Koponen, T., Räsänen, P., Aunola, K., Lerkkanen, M.-K., and Nurmi, J.-E. (2014). Linguistic and spatial skills predict early arithmetic development via counting sequence knowledge. *Child Dev.* 85, 1091–1107. doi: 10.1111/cdev.12173

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

Copyright © 2018 Salminen, Koponen and Tolvanen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Does Multi-Component Strategy Training Improve Calculation Fluency Among Poor Performing Elementary School Children?

Tuire K. Koponen^{1,2*}, Riikka Sorvo³, Ann Dowker⁴, Eija Räikkönen¹, Helena Viholainen³, Mikko Aro³ and Tuija Aro^{2,5}

¹ Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland, ² Niilo Mäki Instituutti, Jyväskylä, Finland, ³ Department of Education, University of Jyväskylä, Jyväskylä, Finland, ⁴ Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom, ⁵ Department of Psychology, University of Jyväskylä, Jyväskylä, Finland

OPEN ACCESS

Edited by:

Emily Kate Farran,
UCL Institute of Education,
United Kingdom

Reviewed by:

Noelia Sánchez-Pérez,
Universidad de Murcia, Spain
Xiaoyi Hu,
Beijing Normal University, China

*Correspondence:

Tuire K. Koponen
tuire.k.koponen@jyu.fi

Specialty section:

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

Received: 09 February 2018

Accepted: 20 June 2018

Published: 11 July 2018

Citation:

Koponen TK, Sorvo R, Dowker A, Räikkönen E, Viholainen H, Aro M and Aro T (2018) Does Multi-Component Strategy Training Improve Calculation Fluency Among Poor Performing Elementary School Children? *Front. Psychol.* 9:1187. doi: 10.3389/fpsyg.2018.01187

The aim of the present study was to extend the previous intervention research in math by examining whether elementary school children with poor calculation fluency benefit from strategy training focusing on derived fact strategies and following an integrative framework, i.e., integrating factual, conceptual, and procedural arithmetic knowledge. It was also examined what kind of changes can be found in frequency of using different strategies. A quasi-experimental design was applied, and the study was carried out within the context of the school and its schedules and resources. Twenty schools in Finland volunteered to participate, and 1376 children were screened in for calculation fluency problems. Children from second to fourth grades were recruited for the math intervention study. Children with low performance (below the 20th percentile) were selected for individual assessment, and indications of using counting-based strategies were the inclusion criteria. Altogether, 69 children participated in calculation training for 12 weeks. Children participated in a group based strategy training twice a week for 45 min. In addition, they had two short weekly sessions for practicing basic addition skills. Along with pre- and post-intervention assessments, a 5-month follow-up assessment was conducted to exam the long-term effects of the intervention. The results showed that children with dysfluent calculation skills participating in the intervention improved significantly in their addition fluency during the intervention period, showing greater positive change than business-as-usual or reading intervention controls. They also maintained the reached fluency level during the 5-month follow-up but did not continue to develop in addition fluency after the end of the intensive training program. There was an increase in fact retrieval and derived fact/decomposition as the preferred strategies in math intervention children and a decrease of the use of counting-based strategies, which were the most common strategies for them before the intervention. No transfer effect was found for subtraction fluency.

Keywords: intervention, calculation fluency, calculation strategies, derived fact, mathematical learning difficulties

INTRODUCTION

Arithmetic calculation is a basic academic skill that, along with reading and writing skills, forms the foundation for academic learning and practical skills of daily life. While there are some national and cultural differences, some studies suggest that approximately 20% of people struggle with basic numerical skills (e.g., Bynner and Parsons, 1997). Studies in several countries suggest that about 5–7% of the population have severe specific mathematical learning difficulties (MD) (Shalev et al., 2005; Butterworth et al., 2011; Geary et al., 2012), although the figure depends on the exact criteria used for diagnosing MD (Kaufmann et al., 2013). In general, the term Mathematical Learning Difficulty (MD) is used broadly to describe a wide variety of deficits in math skills, such as problems in the estimation and processing of quantity and in using the mental number line, in transcoding between number words, digits and quantities or problems in understanding the Base-10 number system or fluently solve simple arithmetic problems and instead use immature counting strategies (e.g., Gersten et al., 2005). It has been proposed that arithmetical fact retrieval deficit resistant to instructional intervention might be a useful diagnostic indicator of arithmetical forms of MLD (Geary, 2004). In the present study we will focus on these arithmetic dysfluency problems.

Difficulties in arithmetic can have serious long-term consequences for later school achievement and limit one's societal and occupational opportunities in adult life (Bynner and Parsons, 1997; Parsons and Bynner, 2005). Individuals with numeracy difficulties tend to leave school early, frequently without qualifications, and have more difficulty than those without such difficulties in getting and maintaining full-time employment (Bynner and Parsons, 1997; Parsons and Bynner, 2005). Gross et al. (2009) estimated that mathematics learning problems reduce an individual's earnings by at least 10%, even after controlling for socio-economic status and other factors. Effective tools for support should be available at schools to provide adequate basic skills and to diminish later difficulties in basic mathematical skills, and thus, prevent long-lasting negative impacts.

Dysfluency in arithmetic calculation, i.e., difficulty in fact retrieval is the most typical feature of MDs. Children with dysfluency problems often rely on slow and error-prone counting strategies, such as *counting all* or *counting on from the first number* (Geary, 2004). They show problems in shifting from immature counting strategies to more advanced strategies, such as direct and fast *fact retrieval*, *decomposing* the problem into smaller facts ($7 + 6 \rightarrow 7 + 3 = 10$, $10 + 3 = 13$), or *deriving* unknown arithmetical facts from known facts ($7 + 6 \rightarrow 6 + 6 = 12 \rightarrow 7 + 6 = 12 + 1 = 13$), despite several years of formal schooling. The differences in math performance between typically performing children and children with MDs can be striking. Even young primary school children can often retrieve answers from memory or derive and predict unknown arithmetical facts from known facts without direct teaching (Dowker, 1998, 2014; Canobi, 2005), whereas children with difficulties may not learn to use these more advanced strategies despite practicing arithmetic at school for several years and

despite having a normal cognitive capacity. Previous intervention research aimed at enhancing calculation fluency in children with MDs has generally focused either on training fact retrieval itself or more efficient counting-based strategies, such as counting on from the largest number, (e.g., Christensen and Gerber, 1990; Tournaki, 2003; Fuchs et al., 2006), and thus the effectiveness of training MD children in derived fact and decomposition strategies remains unclear.

The development of calculation fluency is a multidimensional process. According to the *overlapping waves theory* (Siegler, 1996), one dimension influencing the development of calculation fluency is the frequency of using different strategies: during the typical development of calculation, more efficient strategies (such as fact retrieval and deriving/decomposing) become more dominant. According to this view, difficulties in calculation fluency involve the infrequent use of efficient calculation strategies and the frequent use of slow and error-prone counting-based strategies. Difficulties in calculation fluency and in making the shift to more frequent use of more efficient strategies can stem from several sources. First, it has been suggested that rapid *access to long-term memory* is central for the ability to retrieve arithmetical facts from memory, and that difficulties in this area constitute the key deficit underlying calculation dysfluency among children with MDs, making it difficult for them to use the most efficient strategies. This deficit is particularly marked regarding learning multiplication tables, which is the arithmetical operation mostly relying on arithmetical fact retrieval, and it is also required for fluent addition and subtraction.

The second key deficit might be related to *conceptual knowledge*, which enables individuals to determine the answer to an unknown problem using some known fact, i.e., using derived fact strategies and/or dividing the problem into smaller sums that are easier to solve or retrieve (decomposition), and thus can provide effective back-up strategies when fact retrieval is not possible. Dowker (2009) has suggested that use of these derived fact and decomposition strategies might be an indication of the extent to which children have an explicit understanding of the connections between individual number facts and/or between different arithmetical operations. Thus, a lack of conceptual understanding might be one reason children with MDs do not typically use the more advanced strategies but rely mostly on slower counting-based strategies, such as to start counting from the first addend in the problem (*COF/Counting on from the first number*) rather than the more sophisticated strategy, i.e., *Counting min strategy*, where counting starts from the larger addend. The third deficit is related to the mastery of rules and calculation procedures, i.e., *Procedural knowledge* (Geary, 1993), which means knowing how to use certain arithmetic strategies, such as “borrowing; in subtraction.”

This classification of deficits is in line with the theory that arithmetical knowledge consists of at least three different types of knowledge: factual, conceptual, and procedural (Girelli et al., 2002). Deficit in one type of knowledge might be compensated when using other knowledge as well as by learning how to integrate these knowledge. Difficulty in retrieving arithmetical facts from memory is one of the most consistent findings in the MD literature (e.g., Geary, 1993; Jordan et al., 2003;

Cirino et al., 2007; Geary et al., 2007), and this difficulty is known to be rather persistent. Thus, including two other components, procedural and conceptual knowledge, for training, in addition to fact retrieval, could contribute to the development of compensatory mechanisms for children with difficulties in arithmetical calculation (Girelli et al., 2002).

In recent times, a wide variety of educational interventions have been developed for helping children with difficulties of varying severity in mathematics (Chodura et al., 2015; Dowker, 2017). They have targeted a wide variety of components and subcomponents of arithmetic and have been flexibly adapted to individual children, e.g., Catch Up Numeracy TM (Dowker and Sigley, 2010; Holmes and Dowker, 2013) and Numbers Count (Torgerson et al., 2011). However, it is still true to say that most educational interventions thus far have targeted just one component, most commonly factual knowledge trained by drilling (e.g., Christensen and Gerber, 1990; Hasselbring et al., 1988; Fuchs et al., 2006). Some more recent studies have, however, compared two or more interventions focusing on different components, and thus using different methods of training, e.g., drilling (factual knowledge) with procedural strategy training (with or without conceptual knowledge) versus more general procedural training with multi-digit numbers (Powell et al., 2009), or drilling alone versus a procedural strategy training alone versus the combination of approaches (Woodward, 2006; Fuchs et al., 2008; for review, see Fuchs et al., 2010).

Findings as to the effectiveness of these approaches are mixed. Some of the studies suggest that children with MDs benefit more from strategy instruction or a combination of strategy training and drilling instead of instruction through pure drill and practice, which targets only factual knowledge (Tournaki, 2003). In the study by Tournaki (2003), single-digit addition facts were taught through strategy instruction as well as drill and practice. The results showed that second graders with MDs benefited more from strategy instruction than from instruction through drill and practice (Tournaki, 2003), whereas typically developing controls improved significantly both in the strategy and the drill-and-practice conditions compared to the control condition. However, these two intervention conditions also differed regarding feedback, in that immediate feedback was provided in the strategy condition and delayed feedback in drill-and-practice conditions, so that is difficult to separate the effects of the differences in feedback from those in training (see Powell et al., 2009). In an intervention study by Woodward (2006) with a group of fourth graders (9–10 year olds) with a wide ability range in arithmetic, a combination of strategy training and drilling on facts led to greater improvement in calculation fluency than drilling alone. In contrast, Powell et al. (2009) did not find any differences in post-test performance between children with MDs who received just fact retrieval training and those who received a combination of fact retrieval and strategy instruction among children with MDs. Both intervention groups performed significantly better at the post-test than a business-as-usual control group.

The concept of integrating all three kinds of arithmetical knowledge has been applied in few single-case intervention studies. Case studies with adult (Girelli et al., 2002) and with

child (Koponen et al., 2009) have suggested that if arithmetical fact retrieval is severely impaired and resistant to intervention, a better way of improving children's calculation skills might be to train them in more efficient calculation strategies that rely on procedural and conceptual knowledge. The main principle of both studies was that rather than training children in arithmetical facts by rote learning, the aim was to enable them to use conceptual and procedural knowledge to construct calculation strategies based on meaningful relationships between the known and unknown arithmetical facts. This is important, both because derived fact strategies are themselves an important aspect of arithmetical reasoning (Dowker, 1998, 2014; Canobi, 2005; Star and Rittle-Johnson, 2008) and because children with difficulties in fact retrieval may be able to use such strategies to compensate. Although rigid counting-based strategy use characterizes many children with MDs, some studies suggest that the ability to use derived fact strategies is a relative strength for some low attainers in arithmetic (Russell and Ginsburg, 1984; Dowker, 2009); it may be possible to capitalize on this in enabling them to develop and use compensatory strategies. In Koponen et al. (2009), single-case study, a child was trained to use known arithmetical facts to derive other facts by comparing the magnitude of numbers presented in one arithmetical problem to those of the other problem. He was enabled to determine, based on this comparison, and his previous knowledge of arithmetical operations and principles, how the answers of the two arithmetical problems differed in magnitude (e.g., $5 + 5 = 10$, $5 + 6 = ?$). The procedural training that he received was linked to his existing conceptual knowledge of numbers and arithmetical operations as well as some familiar arithmetical facts, such as $5 + 5 = 10$.

Because there are only a few studies, mostly focusing on single cases, more evidence is needed regarding the effect of strategy intervention integrating the three types of arithmetical knowledge. Besides individually tailored remediation, there is a need for intervention tools and programs that can be effectively applied in small groups or even in classrooms to support calculation fluency among children to whom curriculum-based instruction and training at school is not sufficient to provide adequate calculation skills. This would contribute to such an intervention program becoming sustainable in a school long term, independently of a concurrent research program.

Another gap in the existing intervention literature is that many previous studies focusing specifically on strategies have focused on a rather limited set of strategies emphasizing those usually learned at early phase of typical strategy development. For example, in a study by Tournaki (2003), strategy training included teaching the *minimum addend* strategy, in which the student determines the larger addend and counts on from that cardinal value the number units specified by the smaller addend (e.g., $2 + 6$, students start from 6 and adds two more). Fuchs et al. (2009) carried out an intervention in which children practiced $n + 0$, $n + 1$, $n + 2$ strategies utilizing counting sequence and number knowledge, and although the doubles ($2 + 2$; $6 + 6$ etc.) were trained as well, the focus was on "*know it or count it.*" There have been rather few intervention programs emphasizing alternative calculation strategies, such as derived fact strategies, among children with poor calculation fluency.

There are, however, several studies of interventions involving training in derived fact strategy use, which have tended to yield positive short-term results, but most such studies have either been embedded in practice rather than research and have, for example, lacked control groups (Thornton, 1978; Steinberg, 1985; Adetula, 1996; Askew et al., 2001) or have included derived fact strategy training as just one of many components of an intervention program (Dowker and Sigley, 2010; Holmes and Dowker, 2013; Bakker et al., 2015), making it hard to assess the specific impact of derived fact strategy training.

One study that did compare derived fact strategy training with procedural training was carried out by Caviola et al. (2016). They divided 219 third and fifth graders into three approximately equal groups: a computer-based (derived fact) strategic training group in mental addition, a procedural training group in mental addition, and a business-as-usual control group. Both forms of training had positive effects on addition post-tests, with the strategic training being more effective with the third graders, and the procedural training with the fifth graders. This study did not, however, focus on children with MDs.

Moreover, in previous studies calculation outcome measures have mainly involved calculation fluency and accuracy, not the frequency of use in different strategies. Thus, it does not allow for concluding which type of intervention promotes the use of which strategies (Fuchs et al., 2010). Finally, most of the above-mentioned intervention studies have not examined whether the intervention effect is maintained over time, i.e., whether training enhances the learning only temporarily or whether there are long-term benefits.

Present Study

The present study extends the previous intervention research in math by examining whether children with poor calculation fluency benefit from derived fact strategy training based on an integrative framework (i.e., integrating factual, conceptual, and procedural knowledge training) administered at a school setting in small groups. The long-term benefits of the intervention were assessed 5 months after the intervention ended. The development of the Math intervention group was compared with two different kinds of control groups, one receiving similar kinds of intensive support provided by a special education teacher and implemented in small groups but in a different context (reading intervention group). Another control group consisted of classmates, who were performing the “next poorest” in the classroom, matched for gender (if possible) and who had the same classroom teacher as the Math intervention group and received business-as-usual instruction at school. Both calculation fluency and changes in the frequency of using different kinds of strategies were assessed. The specific research questions were:

- (1) Does explicit strategy training integrating factual, conceptual, and procedural knowledge improve the calculation fluency of addition among children with poor calculation fluency?
- (2) Does the calculation fluency development of the explicit strategy training group differ from that of a control group receiving a similarly intensive reading intervention

(controlling for additional instructional attention and peer group support) or from the development of business-as-usual classmate controls with low performance in calculation fluency?

- (3) Does the explicit strategy training integrating factual, conceptual, and procedural knowledge also change the frequency of use in different strategies?

MATERIALS AND METHODS

Participants

This study was part of a longitudinal *Self-efficacy and Learning Disability Intervention* research project (SELDI; 2013–2015) focusing on elementary school children’s self-beliefs, motivation, and reading and math fluency skills, and in support of children with reading or math difficulties. The data for the present study were collected between November 2013 and October 2014. A total of 20 schools in urban and semi-urban areas in Central and Eastern Finland volunteered to participate, from which the classes and children were recruited for this study. Written consent was obtained from the guardians of the participants. The research procedure was evaluated by the University of Jyväskylä Ethical Committee.

The original sample consisted of 1,327 children (638 girls, 689 boys) from grades 2 to 5. Of the participants, 178 (13.41% of the original sample) were second graders ($M_{\text{age}} = 8.35$ years, $SD = 0.32$ years), 471 (35.49%) were third graders ($M_{\text{age}} = 9.34$ years; $SD = 0.31$ years), 383 (28.86%) were fourth-graders ($M_{\text{age}} = 10.40$ years; $SD = 0.35$ years), and 295 (22.23%) were fifth graders ($M_{\text{age}} = 11.39$ years; $SD = 0.36$ years). A calculation strategy training was provided for children from second to fourth grades.

A quasi-experimental design was applied, as the school, classes, and teachers volunteered to participate, written consent from parents was required to participate, and the study was carried out within the context of the school and its schedules and resources. Screening was conducted according to both reading and calculation fluency, and volunteer teachers were randomized to have either reading or arithmetic training group with or without specific self-efficacy feedback. Approximately half of the children participating in the Math intervention received self-efficacy feedback, following the intervention manual, and the other half received the usual feedback given by special education teacher also providing the strategy training. Both groups had identical strategy training. These two intervention groups were balanced according to the calculation fluency in the pretest. The two groups neither differed in addition fluency at any assessment point nor in development ($p < 0.05$) and were thus treated as a unitary group in the present study.

Screening Procedure for Intervention

Screening for the calculation strategy intervention included two steps. First, all participants from the original sample were assessed in terms of their calculation fluency using group-administered timed calculation tasks (Koponen and Mononen, 2010a, unpublished). Children from grades 2 to 4 whose

performance was at or below the 20th percentile in the calculation fluency task were then selected for individual assessment. Individual assessment included 20 single-digit addition items ($2 + 8$, $5 + 4$, $9 + 6$, $7 + 3$ etc.) presented one by one in a game-like situation. Children were asked to respond as quickly as possible to each item. Only for correct responses given within 3 s, a point was scored. Inclusion criteria for the intervention were that children showed dysfluency, both in the group-administered calculation fluency task (i.e., performance at or below the 20th percentile) and in the individual assessment situation requiring fast fact retrieval or the efficient use of back-up strategies (slow or incorrect response at least 30% of the simple addition items). Altogether, 69 children met this selection criteria and were included in the present analyses. An additional six children with low calculation fluency, but who did not meet the selection criterion, participated in the Math intervention for practical reasons (i.e., to be able to form a group) but were not included in the analyses.

Control Groups

In the present study, the development of the Math intervention group during the baseline, intervention, and follow-up periods was contrasted with the development of the reading intervention controls and the classmate controls. To form the classmate control groups ($N = 69$), one child from the class of each participant of the Math intervention was selected based on having the next-lowest addition fluency score.

Classmate controls were matched for gender (when possible), and they received business-as-usual support, including special education usually provided in the school. The reading intervention group consisted of children with reading fluency deficits who received the intervention as part of the SELDI project in small groups during the same period ($N = 85$; for details, see Aro et al., in press).

Intervention Design and Procedure

We applied an intervention design with two pre-, one post-, and one follow-up assessment. Pre-intervention assessments were conducted in November and January. The 12-week-long interventions started in the end of January. A post-intervention assessment was conducted right after the intervention ended in April, and a follow-up assessment 5 months after ending the intervention in the end of September or in the beginning of October. As an exception, the forced fact retrieval and arithmetic fluency tasks were not repeated in January at the second pre-intervention assessment, and strategy use in free-choice condition was assessed at second but not the first pre-intervention assessment.

All calculation fluency tasks together with reading fluency tasks, non-verbal reasoning tasks, self-efficacy and other questionnaires were administered in groups and conducted during three assessment sessions (30–45 min each) at pre1-, post- and follow-up assessments. At the second pre-intervention assessment shortened assessment battery, including addition and subtraction fluency tasks, was administered during one group assessment session. Group assessment was administered before individual assessment at each time point.

Measures

Calculation Fluency Measures

Basic addition and arithmetic fluency were assessed using one individually administered game-like assessment task administered individually, as well as three group paper-and-pencil tests with time limits.

The individual game-like assessment used a *no-choice technique to assess* addition fluency. The children were shown a card with an addition problem on it and were required to answer correctly less than in 3 s to win the card. For the sake of simplicity, we call this test the *forced fact retrieval task* and the outcome variable *fact retrieval ability*, as has been done in several previous studies (Russell and Ginsburg, 1984; Siegler and Shrager, 1984; Jordan and Montani, 1997; Geary et al., 2000, 2012; Jordan et al., 2003). However, at the same time we must accept the fact that other fast back-up strategies are also possible despite the short time allowed for solving the problem, e.g., derived fact strategies. As a *screening and near transfer task* children were given a 2-min group test of addition fluency (Koponen and Mononen, 2010a, unpublished), which consisted of 120 items with addends smaller than 10. As a *far transfer task*, children were given a similar subtraction test (Koponen and Mononen, 2010b, unpublished) consisting of 120 items with answers in the range of 1 to 9 and 2-min time limit. Another *far transfer task* was the three-minute Basic Arithmetic test (Aunola and Räsänen, 2007), which consists of 30 single-digit and multidigit addition, subtraction, division, and multiplication items. In each test, one point was given from all correctly solved items, and the sum score was counted for each test. Correlation between addition, subtraction and arithmetic tasks in original sample varied from 0.74 to 0.85.

Strategy use in a free-choice condition was assessed with 12 addition items in a similar manner as in the forced fluency task with the exception that children were instructed to solve each addition item in a way that is best for them, i.e., the way that will get the correct answer as quickly as possible. The response time was measured, strategy use was observed, and children were asked to describe/show how they calculated if this was unclear. Strategies were classified into four groups. If a child answered correctly within 3 s and without any signs of using counting, the strategy was classified as *fact retrieval*. If a child's response time was over 3 s but no signs of using a counting strategy were observed or reported or the child reported that he/she used 10 pairs, doubles, or some other known arithmetical fact as a help or used a decomposition strategy, the strategy was classified as *derived fact/decomposition*. If the child's response time was 3 s or more and if the child reported or demonstrated the use of counting, the strategy was classified as mental *counting* or *counting aloud*, depending on whether s/he produced number words silently or aloud.

Background Measures

Non-verbal reasoning was assessed in a group situation using Raven's Colored Progressive Matrices (CPM; Raven et al., 1999). The CPM comprises 36 items divided into three sets of 12 (set A, Ab, and B). Within each set, items are ordered in terms of increasing difficulty. Additionally, vocabulary was assessed

individually using the Vocabulary subtest from the Wechsler Intelligence Scale for Children-IV (WISC-IV; Wechsler, 2010) with Finnish normative data. In this task words of increasing difficulty are presented orally, and children are required to define the words. According to test manual Cronbach's alpha for 8–11 years old varied from 0.83 to 0.87. Visuo-spatial skills were assessed using the Block Design subtest of the WISC-IV (Wechsler, 2010). In this test, the individual is presented with identical blocks with surfaces of red, surfaces of white, and surfaces that are half red and half white. Using an increasing number of these blocks, the individual is required to replicate a pattern that the examiner presents to them—first as a physical model, and then as a two-dimensional picture. The number of blocks required to match the presented models increases, and the patterns become increasingly difficult to visually dissect into components. According to test manual Cronbach's alpha for 8 to 11 years old varied from 0.73 to 0.76. The standardized scores of each test are presented in **Table 1**.

Intervention Program

In the present intervention study, a shortened version of the SELKIS intervention program (Koponen et al., 2011) was used. This program focuses on derived fact strategy training and aims at helping children to discover more efficient calculation strategies using their existing knowledge of number sequences, number concepts, and arithmetical facts (conceptual knowledge). Children participated in the Strategy training group sessions twice a week for 45 min. The number of participants in the groups varied between 4 and 6. In addition, they had two short weekly Gaming sessions for practicing basic addition skills by playing math games and got a worksheet for homework including similar kinds of additions practiced during strategy sessions.

Strategy Training Group Sessions

Addition strategies were trained twice a week in group sessions conducted by special education teacher following the intervention manual. The contents and order of strategy training is presented in **Table 2**. Each session started with checking the homework and followed by instruction sessions, exercises, and closing. Each session consisted of one or two, about 10–15 min' long, strategic instruction sessions as well as of short games and exercises. During the instruction teacher modeled and discussed with children about the magnitude relations between numbers and how counting sequence and addition are linked with this knowledge of number relations (two steps forward in counting sequence – number that is two larger – $x + 2$). Moreover, children were instructed to pay attention and compare how arithmetical facts are related according magnitude ($5 + 5$ and $6 + 5$, six is one more than five, six and five makes one more than five and five). These discussions aimed at guiding the children to discover new strategies based on conceptual understanding. Intervention program manual instructed teachers to encourage children to verbalize their thinking and strategies as well as to point out that use of several strategies is possible and each child should find the fastest strategies for him or herself. After instruction sessions children practiced calculation strategies by playing familiar games embedded with arithmetical contents,

such as a board game with doubles and doubles +1, Bingo, card games with ten pairs.

Gaming Sessions

Short game-like practicing sessions were arranged twice a week each lasting about 15 min. The Gaming sessions were organized and instructed by school assistant or classroom teacher who followed the intervention manual. During these sessions children played games that were already introduced during the Strategy training sessions (card games, board games, etc.) and the aim was to provide repetitions in using addition strategies and achieve fluency. After each session children got a marking (sticker or stamp) to their "game chart."

Teacher Training and Fidelity

Before the intervention periods, researchers instructed all participating teachers on how to implement the intervention program and provided them with detailed session-by-session manuals. Two 3-h-long training sessions were organized including the theory of calculation fluency development as well as how to implement intervention in practice using the program manual. After the third intervention session, researchers called to each teacher to ensure that manuals were followed, and main principals of the programs understood. Moreover, two meetings were arranged during the intervention to share experiences and ensure that all the teachers had common understanding of the key points. Teachers also filled a checklist type of diary, marking the completed intervention sessions and noting any exceptions in intervention activities or attendance of participants. There was altogether 128 activities within 24 strategy training sessions (introduction of strategies, games/exercises, starting and closing activities) and the average amount of activities completed by teachers without exceptions (e.g., didn't have time enough) was 97%. The attendance percentage of individual children varied typically from 92 to 100% in a group meaning that in most of the groups one child was not absent more than 2 times out of 24 intervention sessions. However, there were four children that missed 4 out of 24 intervention sessions, one missed 5 and one 7. All these children were included in the analyses.

Data Analyses

The mean values, mean standard scores and standard deviations of the background variables (Age, Raven's CPM, Block Design, Vocabulary) were calculated. The differences between the Math intervention group and two control groups (Reading intervention controls and Business-as-usual controls) were analyzed by means of independent-samples *t*-tests. Gender differences were analyzed using Chi-square tests. The means and standard deviations for calculation fluency measured variables (fact retrieval, addition fluency, subtraction fluency, and arithmetic fluency) were calculated at each assessment point, and the mean differences between the math intervention group and classmate controls were tested using independent-samples *t*-tests. Differences between the Math intervention group and Reading group were analyzed using univariate analysis of covariance (ANCOVA) using age and gender as the covariate.

TABLE 1 | Descriptive statistics of background variables for the math intervention, reading intervention and business as usual controls.

	Math intervention	Reading intervention	Controls
<i>N</i>	69	85	69
Gender (boys%)	48%	66%*	49%
Age (<i>M</i>)	113.51	123.99***	113.21
<i>SD</i>	10.65	11.48	12.5
Raven ^a (<i>M</i>)	8.74	9.04	9.67
<i>SD</i>	3.81	3.27	3.00
Block design ^a	8.65	9.16	NA
<i>SD</i>	3.22	3.15	NA
Vocabulary ^a	7.89	7.65	NA
	2.75	3.34	NA

^aStandard score (Mean = 10, SD = 3). NA, not available. * $p < 0.05$, *** $p < 0.001$.

The intervention effect in the four outcome measures (fact retrieval, addition fluency, subtraction fluency, and arithmetic fluency) was first analyzed in the Math intervention group using univariate ANOVA for repeated measures (repeated-measures ANOVA) with *time* (pre-test1 vs. pre-test2 vs. post-test vs. follow-up) as a within-subject factor. The partial eta-square was calculated as a measure of effect size. In a second analysis, the progress of the Math intervention group was compared with that of the Business-as-usual controls, and *group* was added as a between-subjects factor. Because there were statistically significant differences in age and gender between the Reading and Math intervention groups, age and gender were used as covariates and univariate analysis of covariance for repeated measures (repeated-measures ANCOVA) as the analysis method. Where an interaction effect was found, planned contrasts on pre-, post-, and follow-up tests scores were conducted.

RESULTS

Descriptive Statistics

The means and standard deviations of the age and standardized scores for the CPM, Block Design, and Vocabulary variables as well as percentage of boys in each group are presented in **Table 1**. There were significantly more boys than expected in the Reading intervention group and more girls in the Math intervention group [$\chi^2(1) = 4.83$, $p < 0.05$], and the children in the Reading group were on average older [$t(141.81) = -5.60$, $p < 0.001$]. As expected, there were no gender or age differences between the Math intervention group and the Business-as-usual controls, as the groups were originally matched for gender and grade [$\chi^2(1) = 0.03$, $p > 0.05$; $T(134) = 0.26$, $p > 0.05$]. Analyses showed that the Math intervention group did not differ significantly on the Raven's Matrices test from either Business-as-usual controls [$t(121.87) = -1.83$, $p > 0.05$] or Reading intervention controls [$T(122.23) = -1.78$, $p > 0.05$]. There were no statistically significant differences between the Math and Reading intervention groups on either the Block Design test [$t(146) = -0.96$, $p > 0.05$] or the Vocabulary test

[$t(144.92) = 0.63$, $p > 0.05$] (data were not available for the Business-as-usual controls).

Efficacy of the Intervention Among Children With Dysfluent Calculation Skills

The means and standard deviations of all calculation fluency measures (fact retrieval, addition fluency, subtraction fluency, and arithmetic fluency) for each group at each assessment point (pretest1, pretest2, post-test, follow-up) are presented in **Table 3**. The results of the repeated-measures ANOVAs for the math group are presented in **Table 4**. Statistically significant effects were found for time in all the calculation fluency tests. Calculation fluency showed favorable development among the Math intervention group throughout the entire study period in all four measured sub-skills. The effect sizes ranged from 0.24 to 0.65. The lowest level of improvement was found for subtraction and the highest for the forced fact retrieval and for addition fluency tasks.

Planned contrast was used to analyze the development of calculation fluency in the Math intervention group in more depth. The analysis of the calculation fluency tasks including addition (fact retrieval, addition fluency, and arithmetic fluency) indicated statistically significant development during the intervention period ($p < 0.001$, $\eta_p^2 = 0.76$; $p < 0.001$, $\eta_p^2 = 0.49$, $p < 0.001$, $\eta_p^2 = 0.36$, respectively). In addition, fluency task data were also available from the baseline period (pretest1–pretest2), showing significant improvement ($p < 0.001$, $\eta_p^2 = 0.35$). From the post-test to follow-up, significant improvement was found in arithmetic fluency ($p < 0.05$, $\eta_p^2 = 0.08$), but not in fact retrieval ($p > 0.05$, $\eta_p^2 = 0.05$) or addition fluency ($p > 0.05$, $\eta_p^2 = 0.01$). In subtraction fluency, the greatest improvement was during the follow-up ($p < 0.001$, $\eta_p^2 = 0.19$) after a very small but significant improvement during the intervention ($p < 0.05$, $\eta_p^2 = 0.06$).

Group Differences in Calculation Fluency Development

First, we analyzed the *fact retrieval*, in which the data were available only for the Math and Reading intervention

TABLE 2 | Contents of math intervention.

Session 1:	Starting session
Sessions 2–4:	Rules for adding one or two ($N + 1$ and $N + 2$) as well as commutativity principle for addition ($a + b = b + a$)
Sessions 5–6:	Add to five (5+, 1/2/3/4/5). Decomposing numbers 6–9 to 5 and x . Verbalizing “five and two makes seven.”
Sessions 7–10:	10-pairs and 10-pairs plus 1 (deriving answer by using 10-pairs, $5 + 6 \rightarrow 5 + 5 = 10$ and $5 + 6$ “is one more” 11);
Session 11:	Add to 10 (structure of numbers 11–19). Verbalizing “10 and 2 makes 12.”
Session 12:	Rehearsal
Sessions 13–14:	Use the structure of five when solving sums with numbers from 5 to 9 ($6 + 7 = 5 + 1 + 5 + 2 = 10 + 3$)
Sessions 15–18:	Doubles and doubles plus 1 (deriving answer by using doubles, $7 + 6 \rightarrow 6 + 6 = 12$ and $7 + 6$ “is one more” 13)
Sessions 19–22:	Add to 9 or 8 (deriving answer by using sums including number 10; $10 + 7 = 17$, $9 + 7$ is one less and $8 + 7$ is two less)
Session 23:	Rehearsal
Session 24:	Ending session

TABLE 3 | Performance at pretest, post-test and follow-up scores and mean differences.

Group	Math ($N = 69$)		R ($N = 85$)		C ($N = 69$)		Paired comparison	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	Math vs. C	Math vs. R
Fact retrieval								
Pre	8.59	4.31	15.73	3.40	NA	NA		Math < R; $F(1,139) = 73.66^{***a}$
Post	15.59	2.80	17.05	2.81	NA	NA		Math < R; $F(1,138) = 3.14$ ns ^a
Follow-up	14.82	3.36	17.03	3.41	NA	NA		Math < R; $F(1,132) = 7.10^{**a}$
Addition								
Pre1	15.84	5.56	29.96	11.28	20.16	5.00	Math < C; $t(136) = -4.82^{***}$	Math < R; $F(1,141) = 39.25^{***a}$
Pre2	19.30	5.67	31.04	11.94	23.07	5.61	Math < C; $t(126) = -3.78^{***}$	Math < R; $F(1,137) = 21.37^{***a}$
Post	26.40	8.86	36.01	14.70	27.32	7.26	Math = C; $t(126) = -0.64$ ns	Math < R; $F(1,140) = 4.13^{*a}$
Follow-up	26.13	8.68	39.40	14.60	30.53	9.02	Math < C; $t(126) = -2.81^{**}$	Math < R; $F(1,135) = 16.22^{***a}$
Subtraction								
Pre1	14.01	5.54	25.52	9.99	18.39	6.34	Math < C; $t(136) = -4.32^{***}$	Math < R; $F(1,142) = 33.86^{***a}$
Pre2	14.76	6.12	23.27	10.10	17.41	6.62	Math < C; $t(127) = -2.36^{*}$	Math < R; $F(1,137) = 15.39^{***a}$
Post	16.15	5.64	26.99	10.78	21.97	6.85	Math < C; $t(126) = -5.28^{***}$	Math < R; $F(1,140) = 25.56^{***a}$
Follow-up	18.88	6.88	30.99	12.70	24.07	8.40	Math < C; $t(126) = -3.84^{***}$	Math < R; $F(1,135) = 22.45^{***a}$
Arithmetic								
Pre	8.13	3.59	13.33	4.08	10.71	4.71	Math < C; $t(125.12) = -4.59^{***}$	Math < R; $F(1,139) = 30.29^{***a}$
Post	10.50	4.30	13.85	5.12	12.68	3.72	Math < C; $t(126) = -3.05^{**}$	Math < R; $F(1,140) = 5.27^{*a}$
Follow-up	11.82	4.03	15.24	4.29	13.98	4.12	Math < C; $t(126) = -3.00^{**}$	Math < R; $F(1,139) = 30.29^{***a}$

Math, math intervention group; R, reading intervention group; C, business-as-usual controls. ^aage and gender as covariate. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE 4 | Within-Group effect among math intervention group in calculation fluency across the time periods and task.

Task	WS effects		WS contrast		
			Baseline	Intervention	Follow-up
Forced fact retrieval	<i>F</i> (<i>df</i>)	118.8 (2, 126)	NA	194.6 (1, 63)	3.5 (1, 63)
	<i>sig.</i>	0.00	NA	0.00	0.07
	η_p^2	0.65	NA	0.76	0.05
Addition fluency (2 min)	<i>F</i> (<i>df</i>)	64.1 (2, 119.5) ^a	33.0 (1.61)	58.0 (1, 61)	0.4 (1, 61)
	<i>sig.</i>	0.00	0.00	0.00	0.56
	η_p^2	0.51	0.35	0.49	0.01
Subtraction fluency (2 min)	<i>F</i> (<i>df</i>)	20.0 (2.6, 162.7) ^b	4.23 (1.62)	4.1 (1, 62)	14.7 (1, 62)
	<i>sig.</i>	0.00	0.04	0.05	0.00
	η_p^2	0.24	0.06	0.06	0.19
Arithmetic fluency (3 min)	<i>F</i> (<i>df</i>)	34.2 (2, 128)	NA	35.3 (1, 64)	5.9 (1, 64)
	<i>sig.</i>	0.00	NA	0.00	0.02
	η_p^2	0.35	NA	0.36	0.08

^aGreenhouse-Geisser; ^bHuynh-Feldt.

TABLE 5 | Within-group and between-group effects among math intervention and control groups in calculation fluency across the time periods and tasks.

Math and reading groups		WS effects		WS contrast Time*Group		
		Time	Time*Group	Baseline	Intervention	Follow-up
Forced fact retrieval	<i>F (df)</i>	3.97(1.82, 235.1) ^{b,c}	31.84 (1.7, 235.1) ^b	NA	55.84 (1, 127)	1.54 (1, 127)
	<i>sig.</i>	0.02	$p < 0.001$	NA	$p < 0.001$	0.22
	η_p^2	0.03	0.20	NA	0.31	0.01
Addition fluency (2 min)	<i>F (df)</i>	1.83 (2.66, 343.52) ^{b,c}	4.03 (2.66, 343.52) ^b	0.93 (1, 129)	7.04 (1, 129)	8.71 (1, 129)
	<i>sig.</i>	0.14	0.01	0.34	0.01	0.01
	η_p^2	0.01	0.03	0.01	0.05	0.06
Arithmetic fluency (3 min)	<i>F (df)</i>	1.05 (2, 264) ^c	5.16 (2, 264)	NA	8.96 (1, 132)	0.14 (1, 143)
	<i>sig.</i>	0.35	0.01	NA	0.01	0.71
	η_p^2	0.01	0.04	NA	0.06	0.01

Math and classmates groups		WS effects		WS contrast Time*Group		
		Time	Time*Group	Baseline	Intervention	Follow-up
Addition fluency (2 min)	<i>F (df)</i>	93.07 (2.1, 236.1) ^a	4.28 (2.1, 236.1) ^a	2.92 (1, 112)	1.0 (1, 112)	1.0 (1, 112)
	<i>sig.</i>	0.00	0.01	0.09	0.04	0.00
	η_p^2	0.45	0.04	0.03	0.04	0.11
Arithmetic fluency (3 min)	<i>F (df)</i>	57.64 (2, 234)	0.29 (2, 234)			
	<i>sig.</i>	.00	0.75			
	η_p^2	0.33	0.00			

^aGreenhouse-Geisser; ^bHuynh-Feldt; ^cAge and gender as covariate.

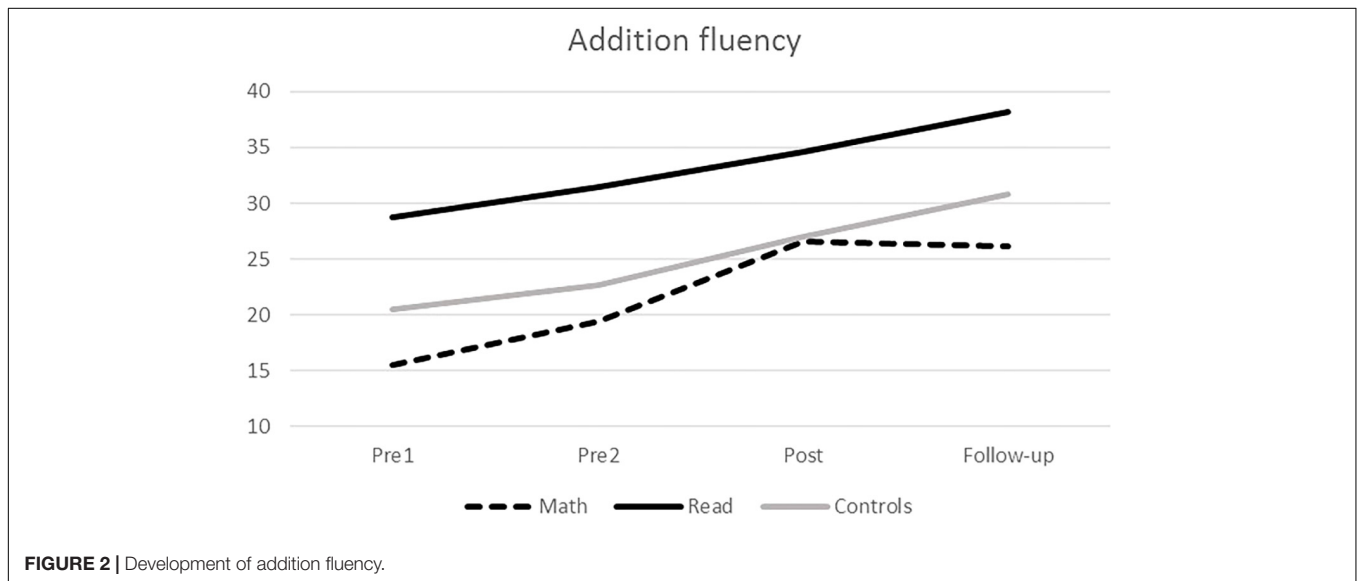
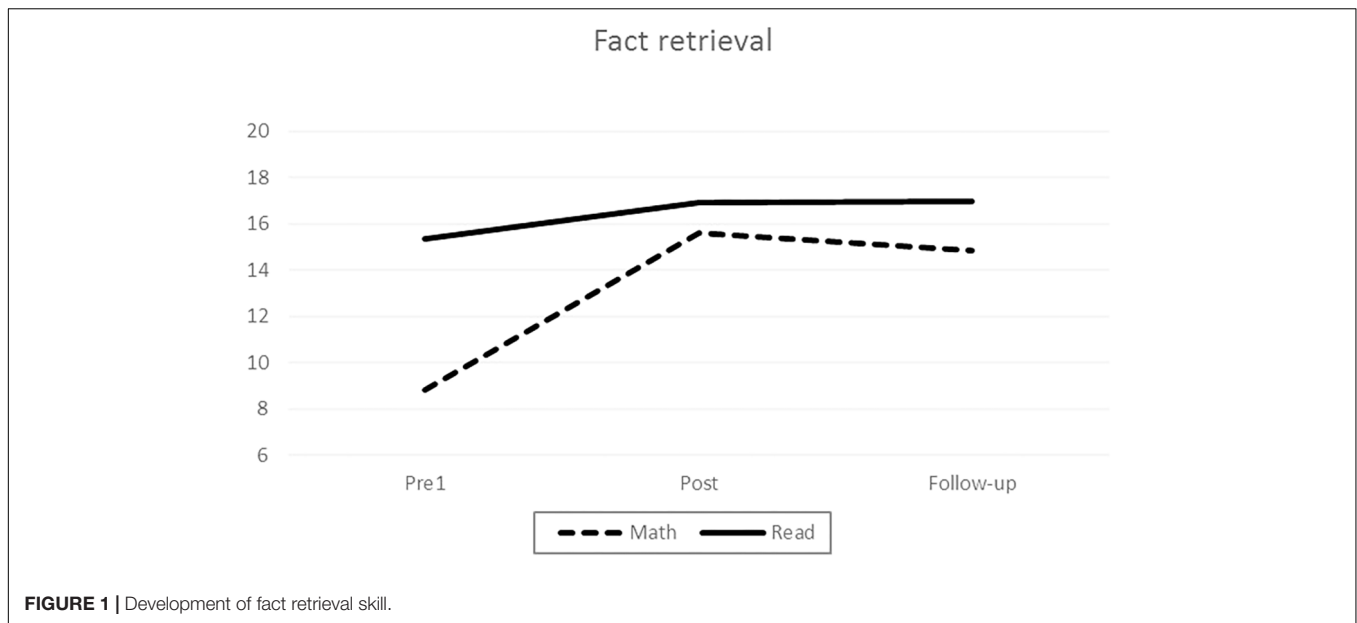
groups. Repeated-measures ANCOVA Statistically significant main effects of time were found, indicating that performance improved in both groups throughout the entire study period (Tables 3, 5). More importantly, there was an interaction between time and group, which was further explored with planned contrasts.

The group-time interaction was statistically significant during the intervention ($p < 0.001$, $\eta_p^2 = 0.31$) but not during the follow-up ($p > 0.05$). As seen in Figure 1, the calculation fluency of the Math intervention group developed clearly during the intervention period, and their skill level remained the same during the follow-up, whereas among the Reading intervention group, small and stable improvements in calculation fluency were found throughout the entire period.

Second, we analyzed the *addition fluency task* using repeated-measures ANCOVA when comparing math intervention group with the reading group and ANOVA with the Business-as-usual controls. When comparing Math intervention group with Reading intervention group non-significant effect for time, gender and age were found. When comparing Math intervention group with Business-as-usual controls, there was a statistically significant main effect of time indicating that performance improved among Math intervention and Business-as-usual groups during the study period (Tables 3, 5). More importantly, there was an interaction between time and group in both analyses as well, which were further explored with planned contrasts. In both analyses, the group $\times$ time interaction was statistically significant during the intervention and follow-up but not during the baseline. As seen in Figure 2, during the intervention

period (i.e., from pre-test2 to post-test) the development of the skills of the three groups differed: although all three groups showed improvement in their skills, the Math intervention group improved faster than the other two groups and did not differ from the Business-as-usual controls at post-intervention assessment (Table 3). At follow-up, the fluency level remained the same in the Math intervention group, while it improved somewhat in the Reading intervention and Business-as-usual control groups. The latter groups showed a constant rate of improvement throughout the study period, while the Math intervention group showed the greatest rate of improvement during the intervention itself.

Third, we analyzed the *arithmetic fluency task* using repeated-measures ANCOVA/ANOVA. When comparing Math intervention group with Reading intervention group non-significant effects for time, gender and age were found but significant interaction between time and group was found, suggesting that the development of arithmetic fluency differed between the Math and Reading groups. This was further explored with planned contrasts. The group $\times$ time interaction was statistically significant during the intervention but not during the follow-up. The Reading and Math intervention groups differed in their level of improvement during the intervention, but not during the follow-up. When comparing Math intervention group with Business-as-usual controls, there was a statistically significant main effect of time indicating that performance improved across Math intervention and Business-as-usual groups during the study period (Tables 3, 5). In contrast, the group $\times$ time interaction was not significant compared to the Math intervention group and Business-as-usual controls. As seen



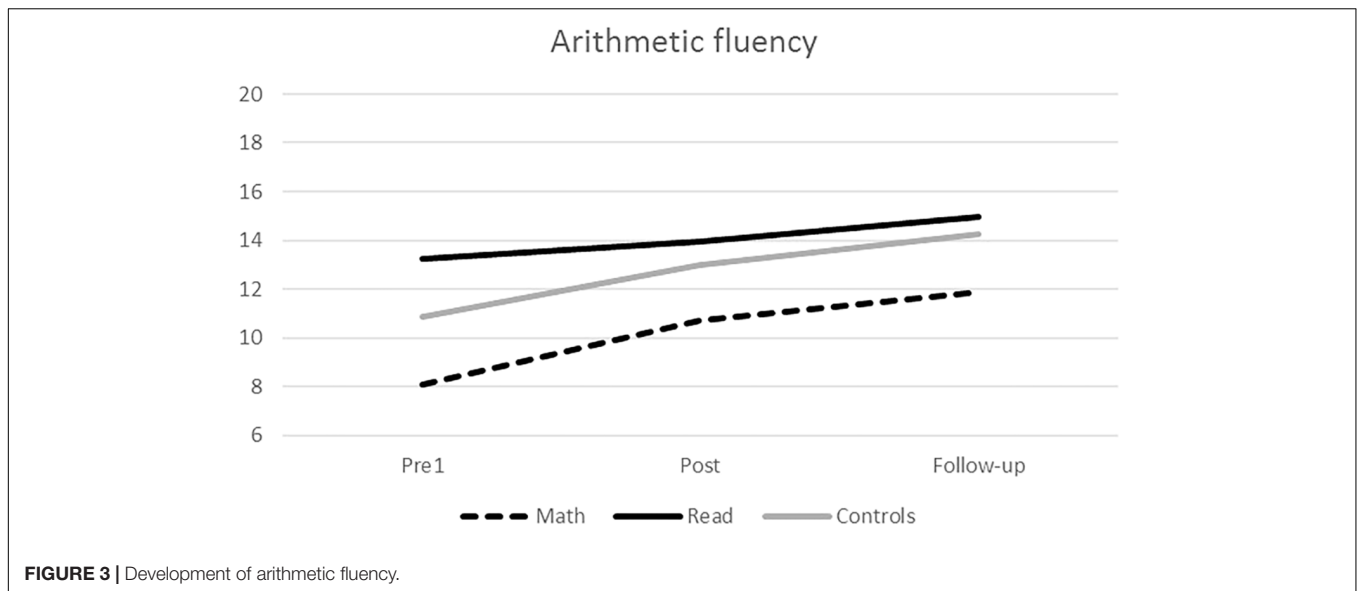
in **Figure 3**, the Math intervention group and the Business-as-usual group showed similar improvement in arithmetic fluency during the intervention period, whereas the Reading intervention group showed slower improvement than the other two groups.

The Math intervention group did not show significant improvement in subtraction fluency during the intervention, and thus further analyses of progress in this subtraction fluency were not carried out.

Changes in the Frequency of Used Strategies in the Free-Choice Condition

Finally, the effects of the explicit strategy training on the frequency of use of different strategies were investigated. As seen in **Figure 4**, before the intervention, *counting in mind*

was the most frequently used calculation strategy among the Math intervention participants, and *fact retrieval* was the most frequently used strategy among the Reading intervention group (*note*: these data were not available for the Business-as-usual controls). After the intervention, fact retrieval became the most frequently used strategy among the Math intervention participants as well; the use of *derived fact strategies* also increased in this group, while the use of *counting-based strategies* decreased among the Math intervention children. All changes during the intervention, the increasing trend in using fact retrieval and derived fact, and the decreasing trend in using counting strategies, were significant among the Math intervention children ($p < 0.05$, $\eta_p^2 = 0.06\text{--}0.44$) but not among the Reading intervention participants ($p > 0.05$) when tested using repeated measures ANCOVA. Using univariate analysis of covariance



(ANCOVA) age and/or gender as the covariate it was further analyzed that the Reading intervention children used significantly more fact retrieval at each time point ($p < 0.05$), although the difference was smaller after the intervention ($\eta_p^2 = 0.03$) than before the intervention ($\eta_p^2 = 0.17$). There were no differences in using deriving strategies before intervention or right after ($p < 0.05$) but the Math group used more deriving strategies after 5 months follow-up ($p < 0.05$, $\eta_p^2 = 0.04$). The Math intervention group used more counting in mind strategies before the intervention ($p < 0.05$, $\eta_p^2 = 0.10$) but statistically significant differences were not found after intervention at post or follow-up assessment ($p > 0.05$). No differences ($p > 0.05$) were found in frequency of using counting aloud strategies at any assessment point, due to infrequent use of this strategy in both groups.

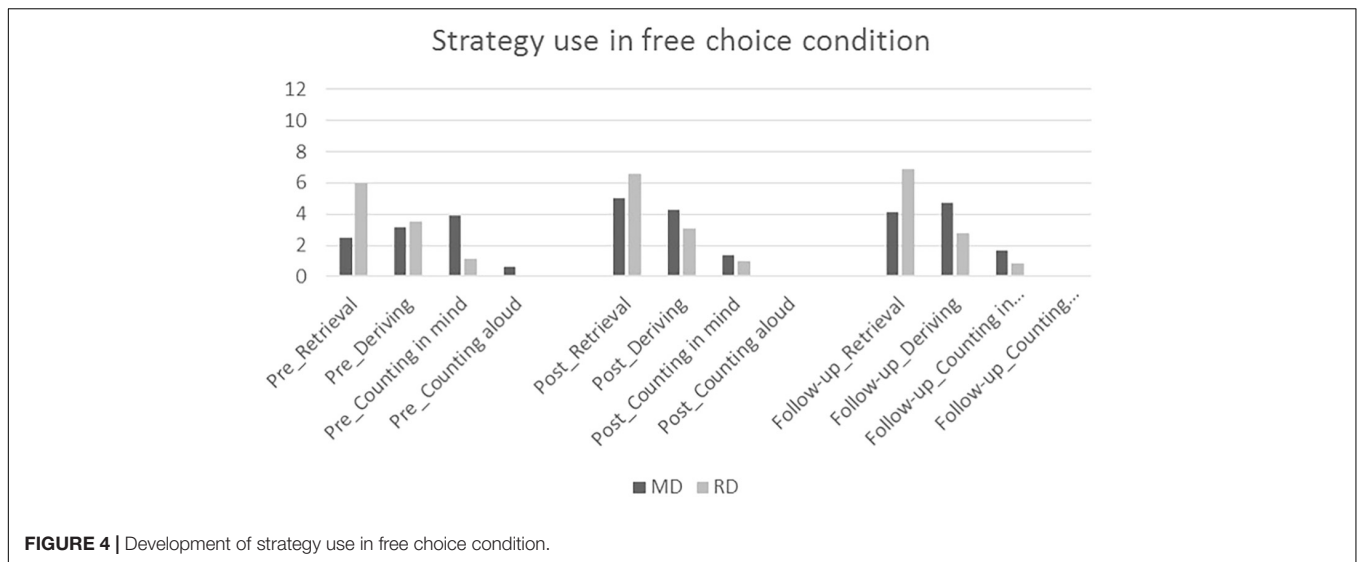
DISCUSSION

The aim of the present study was to extend the previous intervention research in math by examining whether elementary school children with poor calculation fluency benefit from strategy training focusing on derived fact strategies and following an integrative framework (i.e., integrating factual, conceptual, and procedural arithmetic knowledge). The kinds of changes in the frequency of using different strategies were also examined. The SELKIS strategy training program (Koponen et al., 2011) was implemented in small groups by trained special education teachers, highlighting the ecological validity of the present intervention study. Moreover, a 5-month follow-up was conducted to examine the long-term effects of the intervention. The results showed that children with dysfluent calculation skills participating in the Math intervention developed significantly in their addition fluency during the intervention period. They also maintained the reached fluency level during the 5-month follow-up but did not continue to further develop in addition fluency after the intensive training program ended. A similar kind of

developmental slope was found both in fact retrieval as well as in addition fluency assessed in a group situation. Arithmetic fluency, covering all four arithmetical operations and both single-digit as well as multi-digit items, also improved significantly during the intervention period. In contrast, little improvement was found in subtraction fluency during the intervention, and a slightly larger but still very limited improvement was found in subtraction during the 5-month follow-up period.

Further support for a significant effect of the intervention on addition fluency comes from comparing the level of improvement in the Math intervention group with that of the two control groups. Significant group interactions were found in the forced fact retrieval task and in the addition fluency task. The Math intervention group showed more rapid improvement during the intervention than the two control groups and reached the level of Business-as-usual controls at post-intervention assessment point in addition fluency. They maintained the achieved fluency level during the 5-month follow-up but, unfortunately, did not continue to increase their calculation fluency after the intensive intervention period ended. The control groups, in contrast, showed a smooth slope of development in addition fluency throughout the period. The Math intervention group also improved significantly in arithmetic fluency during the intervention period and the interaction between time and group was significant, but their progress did not differ from that of the Business-as-usual controls. Interestingly, the Reading intervention group showed less improvement in arithmetic fluency than either of the other groups.

The maintenance of post-intervention level in addition fluency at the 5-month follow-up assessment provided support for the long-term benefits of the training. However, interesting and important question, as well, is why the Math intervention group did not continue to improve their fluency in addition after the intervention ended. There are several possible explanations. It seems that children with poor calculation fluency require explicit instruction as well as intensive training in order to



extend their arithmetical knowledge and improve efficient strategy use. A summer holiday, lasting two and half months, took a place during follow-up period and naturally can be treated as a non-training period. Moreover, instruction in math typically follows periods including different mathematical contents, such as numbers and operation, geometry, and measurement etc., and thus periods focusing other contents than arithmetic, may not provide intensive training for calculation fluency. In further studies, it should be examined whether we could improve the fluency development after intensive training period by providing material and instruction how to support and strength the use of efficient calculation strategies as part of the Business-as usual instruction in a classroom.

Previous intervention research has generally assessed the efficacy of math intervention training on calculation fluency and accuracy level and not analyzed the changes in strategy use (see Fuchs et al., 2010). In the present study, in a free-choice condition math intervention children increased their use of fact retrieval and derived fact/decomposition as preferred strategies and decreased their use of counting-based strategies, which were their most common strategies before the intervention. The Math intervention group differed from the Reading group using more frequently counting-based strategies before the intervention. Differences were not significant after the intervention. Although the differences in use of retrieval strategies was significant in all assessment points favoring the Reading intervention group, the difference was clearly smaller after the intervention than before it. Moreover, the Math intervention participants used more derived fact/decomposition strategies at follow-up assessment than the Reading group. This finding suggests that despite having dysfluency in basic calculation skills after several years of training at school, explicit instruction utilizing an integrative framework in calculation strategy training can help children to use more often efficient back-up strategies and fact retrieval instead of counting.

The finding related to the missing transfer effect of addition strategy training to subtraction fluency was unfortunate. Moreover, in arithmetic fluency tasks, no developmental trend was identified among math intervention children that would have differed from their classmate control; thus, significant development in arithmetic fluency cannot be concluded to result from the intervention but could be due to schooling in general. However, this finding was not highly surprising, considering that a typical feature of children with MDs is a difficulty in spontaneously discovering efficient calculation strategies (Geary, 1993). It is likely that different arithmetical facts and arithmetic operations, e.g., addition and subtraction, are more isolated for MD children and for this reason they cannot use their knowledge of addition facts when solving subtraction problems. This could explain why spontaneous transfer did not happen across the arithmetic operations, although children started to make more frequent use of retrieval and derived fact/decomposition strategies in addition. Moreover, even typically achieving children often fail to extend their knowledge of addition principles appropriately to subtraction principles (Dowker, 1998, 2014). For example, they find the addition/subtraction inverse principle far more difficult to recognize and use than addition-specific principles, such as commutativity, and often overextend addition principles to subtraction, e.g., saying that if $14 - 5 = 9$, $14 - 6$ must be 10 “because 6 is one more than 5.” Thus, explicit instruction and intensive practice are likely to be required to learn to use derived fact/decomposition strategies for subtraction, rather than expecting them to spontaneously extend their strategic knowledge in addition also to subtraction.

Limitations and Further Directions

Some limitations of the study should be considered when interpreting the current findings. The main limitations are related to the quasi-experimental nature of the design. Since the study was conducted in ecologically valid conditions as part of everyday school routines, blind and fully random matching

of the participants was not achievable. However, children were carefully selected for interventions, and participants showed signs of dysfluency both in group-administered addition fluency task (where items were presented as a list) and in individually administered assessments (where items were presented one at the time). The inclusion of individual assessment is an essential strength of the participant selection, as group tests are not optimal for all children to show their abilities and could more likely lead to identifying false-positive cases. The most serious limitation may be that we did not have data from the Business-as-usual controls for all measures. However, the Reading intervention group data were available for all measures, and this is a more stringent control group.

Another limitation is that, due to the limited resources available, procedures that would allow full monitoring of the reliability and validity of the interventions (e.g., video-recordings) could not be implemented. The measures taken to guarantee the fidelity (teacher training, session-by-session manual, filling diary, meetings and phone calls during the intervention) support the understanding that the programs were implemented following the program manual and intervention design.

Finding a significant intervention effect for low-attaining children, which also remained during the follow-up period, is a positive and promising result, but at the same time only the first step. Further studies comparing this kind of integrative framework to other intervention approaches with even longer follow-up and other age groups are needed to clarify the question of what the most efficient intervention approaches for low attaining pupils are. It would also be beneficial to explore whether the intervention is equally effective in all age groups, especially given Caviola et al. (2016) findings on the differential effectiveness of derived fact training and procedural training in the third- and fifth-grade groups.

It would also be desirable to investigate the specificity of the effects, both within arithmetic and between arithmetic and other

subjects. In this study, training in addition had little impact on subtraction. Further research is recommended to determine whether the same would be found regarding the effect of training in subtraction on addition.

Despite the positive findings related to the intervention effect, it should be noted that, as found in other intervention studies, there were differences in responsiveness among intervention participants. In the future, the variation in responsiveness should be studied to better understand the factors influencing the benefits of derived fact strategy training within an integrative framework, and to gain a better understanding of how to target interventions for groups of participants, and to maximize their effectiveness.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of National Advisory Board on Research Ethics. The protocol was approved by the National Advisory Board on Research Ethics. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

TK drafted a first version of the current paper. RS, AD, ER, HV, MA, and TA were responsible for draft editing.

FUNDING

This research was supported by the Academy of Finland (Grant Numbers 264344 and 264415 for 2013–2015 and 292466 for 2015–2019).

REFERENCES

- Adetula, L. O. (1996). Effects of counting and thinking strategies in teaching addition and subtraction problems. *Educ. Res.* 38, 183–198. doi: 10.1080/0013188960380206
- Aro, T., Viholainen, H., Koponen, T., Peura, P., Räikkönen, E., Salmi, P., et al. (in press). Can reading fluency and self-efficacy of reading fluency be enhanced with an intervention targeting the sources of self-efficacy? *Learn. Individ. Differ.*
- Askew, M., Bibby, M., and Brown, M. (2001). *Raising Attainment in Primary Number Sense: From Counting to Strategy*. London: Beam Education.
- Aunola, K., and Räsänen, P. (2007). *The Basic Arithmetic Test*. Jyväskylä: Niilo Mäki Institute and University of Jyväskylä.
- Bakker, M., Van den Heuvel-Oppenhuis, M., and Robzitsch, A. (2015). Effects of mathematics computer games on special education students' multiplicative reasoning ability. *Contemp. Educ. Psychol.* 40, 55–71. doi: 10.1016/j.cedpsych.2014.09.001
- Butterworth, B., Varma, S., and Lorillard, D. (2011). Dyscalculia: from brain to education. *Science* 332, 1049–1053. doi: 10.1126/science.1201536
- Bynner, J., and Parsons, S. (1997). *Does Numeracy Matter?*. London: Basic Skills Agency.
- Canobi, K. H. (2005). Children's profiles of addition and subtraction understanding. *J. Exp. Child Psychol.* 92, 220–246. doi: 10.1016/j.jecp.2005.06.001
- Caviola, S., Gerotto, G., and Mammarella, I. (2016). Computer-based training for improving mental calculation in third- and fifth-graders. *Acta Psychol.* 171, 118–127. doi: 10.1016/j.actpsy.2016.10.005
- Chodura, S., Kuhn, J.-T., and Holling, H. (2015). Interventions for children with mathematical difficulties: a meta-analysis. *Z. Psychol.* 223, 129–144. doi: 10.1027/2151-2604/a000211
- Christensen, C. A., and Gerber, M. M. (1990). Effectiveness of computerized drill and practice games in teaching basic math facts. *Exceptionality* 1, 149–165. doi: 10.1080/09362839009524751
- Cirino, P. T., Ewing-Cobbs, L., Barnes, M., Fuchs, L. S., and Fletcher, J. M. (2007). Cognitive arithmetic differences in learning disability groups and the role of behavioral inattention. *Learn. Disabil. Res. Pract.* 22, 25–35. doi: 10.1111/j.1540-5826.2007.00228.x
- Dowker, A. (1998). "Individual differences in arithmetical development," in *The Development of Mathematical Skills*, ed. C. Donlan (London: Psychology Press), 275–302.
- Dowker, A. (2009). Derived fact strategies in children with and without mathematical difficulties. *Cogn. Dev.* 24, 401–410. doi: 10.1016/j.cogdev.2009.09.005
- Dowker, A. (2014). Young children's use of derived fact strategies for addition and subtraction. *Front. Hum. Neurosci.* 7:924. doi: 10.3389/fnhum.2013.00924

- Dowker, A. (2017). Interventions for primary school children with difficulties in mathematics. *Adv. Child Dev. Behav.* 53, 255–287. doi: 10.1016/bs.acdb.2017.04.004
- Dowker, A., and Sigley, G. (2010). Targeted interventions for children with arithmetical difficulties. *Br. J. Educ. Psychol.* 7, 65–81. doi: 10.1348/97818543370009X1258369932492
- Fuchs, L., Fuchs, D., Hamlet, C. L., Powell, S. R., Capizzi, A. M., and Seethaler, P. M. (2006). The effects of computer-assisted instruction on number combination skill in at-risk first graders. *J. Learn. Disabil.* 39, 467–475. doi: 10.1177/00222194060390050701
- Fuchs, L. S., Powell, S. R., Hamlett, C. L., Fuchs, D., Cirino, P. T., and Fletcher, J. M. (2008). Remediating computational deficits at third grade: a randomized field trial. *J. Res. Educ. Effect.* 1, 2–32. doi: 10.1080/19345740701692449
- Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., et al. (2009). Remediating number combination and word problem deficits among students with mathematics difficulties: a randomized control trial. *J. Educ. Psychol.* 101, 561–576. doi: 10.1037/a0014701
- Fuchs, L. S., Powell, S. R., Seethaler, P. M., Cirino, P. T., Fletcher, J. M., Fuchs, D., et al. (2010). The effects of strategic counting instruction, with and without deliberate practice, on number combination skill among students with mathematics difficulties. *Learn. Individ. Diff.* 20, 89–100. doi: 10.1016/j.lindif.2009.09.003
- Geary, D. C. (1993). Mathematical disabilities: cognitive, neuropsychological, and genetic components. *Psychol. Bull.* 114, 345–362. doi: 10.1037/0033-2909.114.2.345
- Geary, D. C. (2004). Mathematics and learning disabilities. *J. Learn. Disabil.* 37, 4–15. doi: 10.1177/00222194040370010201
- Geary, D. C., Hamson, C. O., and Hoard, M. K. (2000). Numerical and arithmetical cognition: a longitudinal study of process and concept deficits in children with learning disability. *J. Exp. Child Psychol.* 77, 236–263. doi: 10.1006/jecp.2000.2561
- Geary, D. C., Hoard, M. K., and Bailey, D. H. (2012). Fact retrieval deficits in low achieving children and children with mathematical learning disability. *J. Learn. Disabil.* 37, 291–307. doi: 10.1177/0022219410392046
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., and Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Dev.* 78, 1343–1359. doi: 10.1111/j.1467-8624.2007.01069.x
- Gersten, R., Jordan, N. C., and Flojo, J. R. (2005). Early identification and interventions for children with mathematics difficulties. *J. Learn. Disabil.* 28, 293–304. doi: 10.1177/00222194050380040301
- Girelli, L., Bartha, L., and Delazer, M. (2002). Strategic learning in the rehabilitation of semantic knowledge. *Neuropsychol. Rehabil.* 12, 41–61. doi: 10.1080/09602010143000149
- Gross, J., Hudson, C., and Price, D. (2009). *The Long-Term Costs of Numeracy Difficulties*. London: Every Child a Chance Trust (KPMG).
- Hasselbring, T., Goin, L., and Bransford, J. (1988). Developing math automaticity in learning handicapped children: the role of computerized drill and practice. *Focus Except. Children* 20, 1–7. doi: 10.17161/fec.v20i6.7504
- Holmes, W., and Dowker, A. D. (2013). Catch Up Numeracy: a targeted intervention for children who are low attaining in mathematics. *Res. Math. Educ.* 15, 249–265. doi: 10.1080/14794802.2013.803779
- Jordan, N. C., Hanich, L. B., and Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Dev.* 74, 834–850. doi: 10.1111/1467-8624.00571
- Jordan, N. C., and Montani, T. O. (1997). Cognitive arithmetic and problem solving: a comparison of children with specific and general mathematics difficulties. *J. Learn. Disabil.* 30, 624–634. doi: 10.1177/002221949703000606
- Kaufmann, L., Mazzocco, M. M., Dowker, A., Von Aster, M., Göbel, S. M., Grabner, R. H., et al. (2013). Dyscalculia from a developmental and differential perspective. *Front. Psychol.* 4:516. doi: 10.3389/fpsyg.2013.00516
- Koponen, T., Aro, T., and Ahonen, T. (2009). Conceptual knowledge-based strategy training in single-digit calculation: a single case intervention study. *Eur. J. Special Needs Educ.* 24, 259–276. doi: 10.1080/08856250903016813
- Koponen, T., Mononen, R., Kumpulainen, T., and Puura, P. (2011). *SELKIS-Yhteenlaskua Ymmärtämään. Yhteenlaskutaidon Harjoitusohjelma. [Improving Addition Skills: SELKIS Intervention Program]*. Jyväskylä: Niilo Mäki Institute and Haukkarannan koulu.
- Parsons, S., and Bynner, J. (2005). *Does Numeracy Matter More?*. London: NRDC.
- Powell, S. R., Fuchs, L. S., Fuchs, D., Cirino, P., and Fletcher, J. M. (2009). Effects of fact retrieval tutoring on third-grade students with math difficulties with and without reading difficulties. *Learn. Disabil. Res. Pract.* 24, 1–11. doi: 10.1111/j.1540-5826.2008.01272.x
- Raven, J., Court, J. H., and Raven, J. C. (1999). *Manual for Raven's Progressive Matrices and Vocabulary Scales - Section 2: Coloured Progressive Matrices*. Oxford: Oxford Psychologists Press.
- Russell, R., and Ginsburg, H. P. (1984). Cognitive analysis of children's mathematical difficulties. *Cogn. Instr.* 1, 217–244. doi: 10.1207/s1532690xci0102_3
- Shalev, R. S., Manor, O., and Gross-Tsur, V. (2005). Developmental dyscalculia: a prospective six-year follow-up. *Dev. Med. Child Neurol.* 47, 121–125. doi: 10.1111/j.1469-8749.2005.tb01100.x
- Siegler, R. S. (1996). *Emerging Minds: The Process of Change in Children's Thinking*. New York, NY: Oxford University Press.
- Siegler, R. S., and Shrager, J. (1984). "Strategy choice in addition and subtraction: How do children know what to do?" in *Origins of Cognitive Skills*, ed. C. Sophian (Hillsdale, NJ: Erlbaum), 229–293.
- Star, J. R., and Rittle-Johnson, B. (2008). Flexibility in problem solving: the case of equation solving. *Learn. Instr.* 18, 565–579. doi: 10.1063/1.4834638
- Steinberg, R. (1985). Instruction in derived facts strategies in addition and subtraction. *J. Res. Math. Educ.* 16, 337–355. doi: 10.2307/749356
- Thornton, C. A. (1978). Emphasizing thinking strategies in basic fact instruction. *J. Res. Math. Educ.* 9, 214–227. doi: 10.2307/748999
- Torgerson, C. J., Wiggins, A., Torgerson, D. J., Ainsworth, H., Barmby, P., Hewitt, C., et al. (2011). *Every Child Counts: The Independent Evaluation Executive Summary*. London: Department for Education (DfE).
- Tournaki, N. (2003). The differential effects of teaching addition through strategy instruction versus drill and practice to students with and without learning disabilities. *J. Learn. Disabil.* 36, 449–458. doi: 10.1177/00222194030360050601
- Wechsler, D. (2010). *WISC-IV - Wechsler Intelligence Scale for Children—IV*. Helsinki: Psykologien Kustannus.
- Woodward, J. (2006). Developing Automaticity in Multiplication Facts: integrating strategy instruction with timed practice drills. *J. Disabil. Q.* 40, 270–281. doi: 10.2307/30035554

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

Copyright © 2018 Koponen, Sorvo, Dowker, Räikkönen, Viholainen, Aro and Aro. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Children's Strategy Choices on Complex Subtraction Problems: Individual Differences and Developmental Changes

Sara Caviola^{1*}, Irene C. Mammarella^{2*}, Massimiliano Pastore² and Jo-Anne LeFevre³

¹ Department of Psychology, University of Cambridge, Cambridge, United Kingdom, ² Department of Developmental Psychology, Università degli Studi di Padova, Padova, Italy, ³ Department of Psychology, Institute of Cognitive Science, Carleton University, Ottawa, ON, Canada

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

Catherine Thevenot,
Université de Lausanne, Switzerland
Robert Reeve,
University of Melbourne, Australia

*Correspondence:

Sara Caviola
sc2014@cam.ac.uk
Irene C. Mammarella
irene.mammarella@unipd.it

Specialty section:

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

Received: 24 March 2018

Accepted: 25 June 2018

Published: 17 July 2018

Citation:

Caviola S, Mammarella IC, Pastore M
and LeFevre J-A (2018) Children's
Strategy Choices on Complex
Subtraction Problems: Individual
Differences and Developmental
Changes. *Front. Psychol.* 9:1209.
doi: 10.3389/fpsyg.2018.01209

We examined how children's strategy choices in solving complex subtraction problems are related to grade and to variations in problem complexity. In two studies, third- and fifth-grade children ($N \approx 160$ each study) solved multi-digit subtraction problems (e.g., 34–18) and described their solution strategies. In the first experiment, strategy selection was investigated by means of a *free-choice* paradigm, whereas in the second study a *discrete-choice* approach was implemented. In both experiments, analyses of strategy repertoire indicated that third-grade children were more likely to report less-efficient strategies (i.e., counting) and relied more on the right-to-left solution algorithm compared to fifth-grade children who more often used efficient memory-based retrieval and conceptually-based left-to-right (i.e., decomposition) strategies. Nevertheless, all strategies were reported or selected by both older and younger children and strategy use varied with problem complexity and presentation format for both age groups. These results supported the overlapping waves model of strategy development and provide detailed information about patterns of strategy choice on complex subtraction problems.

Keywords: mental calculation, subtraction problems, strategy choice, children, mathematics, problem solving, arithmetic

INTRODUCTION

Understanding how children choose and apply a specific strategy to solve a mathematical problem is an important issue in the field of numerical cognition. Individuals' strategy choices are influenced by many factors, including their repertoire or knowledge of strategies (Kilpatrick et al., 2001; Baroody and Dowker, 2003), their expertise in implementing those strategies (Torbeys et al., 2006; Verschaffel et al., 2007), and their overall level of mathematical achievement (Geary et al., 2000; Smith-Chant and LeFevre, 2003). Although some studies have examined children's performance in solving complex arithmetic problems, the results have varied depending on children's age and consequently their arithmetic expertise, the specific arithmetic operation, and the type of strategy assessment applied (e.g., Torbeys et al., 2009a,b; Torbeys and Verschaffel, 2016; Lemaire and Brun, 2018). Accordingly, in the present research we conducted two large scale studies using two different

types of strategy assessment in which we investigated how children at different stages of their primary education perform complex subtraction problems varying in the degree of complexity. By manipulating these key variables, we were able to provide a comprehensive overview of the factors influencing children's strategy choices on multi-digit subtraction problems.

Among the four basic arithmetic operations, that is, addition, subtraction, multiplication, and division, subtraction is of particular interest because children and adults report using a wide range of different strategies, both on simple problems such as $14-6$ (e.g., LeFevre et al., 2006) and on complex problems such as $24-9$ or $56-23$ (Torbeys et al., 2009a; Linsen et al., 2014). According to Lemaire and Siegler (1995), arithmetical strategy use entails two components: first choosing a strategy (*strategy choice*), and then implementing the chosen strategy (*strategy execution*) to solve the arithmetic problem. In the present study, the focus was on the process of strategy choice. Strategy choice is associated not only with an individual's level of skill in simple arithmetic (Smith-Chant and LeFevre, 2003), but also with the specific features of a given problem (e.g., problem difficulties in terms of number of digits in each operand or the presence/absence of a borrow procedure; Imbo et al., 2007), and the context of the problem, such as the presentation format (e.g., problems presented in horizontal vs. vertical format Trbovich and LeFevre, 2003; Lemaire and Calliès, 2009; Imbo and LeFevre, 2010). Like Lemaire and Brun (2018), in the present study we provide a detailed analysis of strategies used by children to solve subtraction problems, investigating for the first time how different problem features (i.e., problem complexity and presentation format together) influence children's strategies and performance, as well as how such strategic behavior changes with children's age and related expertise.

Subtraction Strategies

Adults and children have been observed to use a variety of strategies on subtraction problems and these strategies can be categorized according to the type of computations involved. For the very simplest problems, such as $5-3$, memory retrieval is usually reported, but various counting strategies, such as counting up (i.e., for $5-3$, counting 4, 5) are also used (e.g., LeFevre et al., 2006). Solvers may also report counting up or down on large problems such as $52-49$ (Torbeys et al., 2009b). Another way of categorizing strategies on more complex problems is to consider the solution path, such that strategies can be divided into two main categories (Green et al., 2007; Imbo and LeFevre, 2009). In *right-to-left* strategies, the operands are treated as concatenations of single digits and calculation considers each columnar operation separately, for example, solving $45-29$ by subtracting $15-9 = 6$ and $3-2 = 1$. For subtraction, the right-to-left strategy is a mental version of column-by-column algorithm taught at school for use in written calculation. In contrast, in *left-to-right* strategies, operands are represented and manipulated in a more holistic manner. For example, $64-12$ can be decomposed into $60-10 = 50$ and $4-2 = 2$, and then reassembled to obtain the answer (i.e., 52; Lemaire and Calliès, 2009), or $69-13$ can be solved by rounding to $70-13 = 57$ and then subtracting 1. Often referred to as transformation or decomposition strategies,

the left-to-right approach requires conceptual understanding of the structure of numbers (LeFevre et al., 2006).

Children show developmental and educational changes in the use of the left-to-right and right-to-left strategies described above, mainly corresponding to strengthening of their mental calculation skills due to the acquisition of multi-digit algorithms (written) procedures (Fuson et al., 1988; Geary et al., 2004). In particular, in European countries such as Belgium, Italy, and the Netherlands, children are taught mental computation for multi-digit subtraction in second grade, and start to learn the written algorithm in third grade (Cornoldi and Lucangeli, 2004; Caviola et al., 2014). However, in the US and Canada, less emphasis is placed on mental computation, especially for multi-digit problems, which are thought to be solved through the written algorithm (Baroody and Dowker, 2003). Accordingly, strategy use might vary across countries and school systems.

A few researchers have studied children's strategies on complex subtraction but none has documented the full range of strategies used (Beishuizen, 1993; Lemaire and Calliès, 2009). For example, Lemaire and Calliès (2009) compared the performance of 20 fifth- and 20 seventh-grade students in France on complex subtractions, however, they restricted children's strategy choices to two left-to-right methods (i.e., full and partial decomposition). Other researchers have focused on subtraction-by-addition strategies (e.g., Linsen et al., 2014). Thus, currently there is no information about the extent of children's spontaneous strategy use on complex subtraction.

Other important factors in children's strategy choices, beyond their levels of automaticity and strategy knowledge, are the characteristics of the problem. Problem complexity is one feature that is assumed to influence children's strategy choices (Lemaire and Lecacheur, 2011; Ardiale and Lemaire, 2013). One way to vary problem complexity is to include a carry or borrow requirement (Noël et al., 2001; Imbo et al., 2007). Superficial features of the problems, such as presenting the problems in a horizontal vs. a vertical format, may also influence children's strategy choice. Overall, processing efficiency varies with presentation format, and complexity (Trbovich and LeFevre, 2003; Lemaire and Calliès, 2009; Imbo and LeFevre, 2010; Caviola et al., 2012), but, to the best of our knowledge, none of the previous studies examined in a larger set of problems how different features can interact to influence children's strategy choice. Moreover, researchers have suggested that manipulations of presentation format can trigger differential recruitment of cognitive resources, leading to variability in the solution procedures that participants select as a function of format (Trbovich and LeFevre, 2003). For example, vertically-presented problems required more visual resources, whereas horizontally-presented problems required more phonological resources (Caviola et al., 2012). Thus, presentation format may influence selection of strategies, however, this possibility has not been assessed in children.

Methods for Studying Strategy Choices

In the strategy literature, among others, the two most widely used methods implemented to evaluate strategy choice and execution are *free-choice* report and *forced-choice* methods. Free-choice

reports require participants to verbally report which strategy they used to solve a problem immediately after solving it with no restrictions on the strategy repertoire (e.g., Siegler, 1988; Davis and Carr, 2001; Mabbott and Bisanz, 2003; Noël et al., 2004; Torbeyns et al., 2005). This method leads to a broader collection of strategies than forced-choice methods, under the assumption that participants have sufficient metacognitive/introspective and verbal abilities enabling them to provide accurate verbal report of strategies used (Kirk and Ashcraft, 2001). Free-choice reports may not be valid for processes that are fast and automatic (Ericsson and Simon, 1993), and may not be appropriate for children if their verbal abilities are limited (Siegler and Stern, 1998). On this view, free-choice reports may potentially lead to biases, such as the over-reporting of strategies whose salience is high, and conversely, the under-reporting of fast or less procedural ones (Ericsson and Simon, 1993; Kirk and Ashcraft, 2001; Thevenot et al., 2010).

To address some of the limitations of free-choice retrospective reports, Siegler and Lemaire (1997) developed a modified method, known as the choice/no-choice paradigm. This method consists of a forced choice condition, where individuals apply a preferred strategy (chosen from a restricted set of two or three options), and two or more no-choice conditions, where they are required to solve all problems with a single strategy; thus, the number of no-choice conditions corresponds to the number of options in the forced-choice condition (e.g., Imbo and Vandierendonck, 2007a,b,c; Reed et al., 2015). The choice/no-choice paradigm provides information on strategy *efficiency* from the no-choice condition, independently of the choice process, whereas the comparison of performance in the no-choice and the choice condition gives an indication of people's strategy *adaptivity* (i.e., the selection of the most efficient procedure from the limited set provided). Although this method does address some of the problems of the free choice approach, specifically, the concern that strategy choice and efficiency are confounded, there are also criticisms (Luwel et al., 2009). In particular, the choice/no-choice approach limits the strategies that are available because a limited number of no-choice conditions are included and thus may not provide the "best" strategy on any given trial (cf. Imbo and LeFevre, 2011; Xu et al., 2014). In order to address these limitations, some authors have implemented a method that we can define as a *discrete-choice method*, where participants were asked to select between a larger number of given strategy alternatives, within the prospective that providing a broader choice of strategies gives better information about children's choices (e.g., Lemaire and Brun, 2018). The strength of this *discrete-choice* approach lies in the opportunity to determine the effects of explanatory variables from a more extensive set of options, which provides less restriction on individuals' strategy selection (Xu et al., 2014). A similar *discrete-choice* strategy method, in which a set of options was provided, has been used extensively within the domain of simple arithmetic with adults (e.g., Campbell and Xue, 2001; Campbell and Austin, 2002; Imbo and Vandierendonck, 2007a, 2008) and with children, combined with a choice/no-choice experimental design (Imbo and Vandierendonck, 2007b).

In the present study, rather than focusing solely on one single method to assess strategy choice, we tested two cohorts of third- and fifth-school graders in two different experiments in order to indirectly compare free-choice and a discrete-choice approaches.

The Present Research

The central questions in the present research were how children's strategy choices are influenced by (a) their level of expertise (e.g., grade 3 vs. 5), (b) problem features (e.g., complexity and contextual features), and (c) the type of strategy assessment to collect strategy reports. In two experiments, children in the same age-range were tested with the same pool of multi-digit subtraction problems (e.g., 23–3; 47–19). Across experiments, two methods of assessing children's strategy choice were used. In Experiment 1, strategies were assessed on a problem-by-problem basis by means of immediate retrospective verbal self-reports. In Experiment 2, children were asked to choose the strategies used from among a list of alternatives that was based on the solution procedures observed in Experiment 1.

In order to assess the role of expertise in strategy choices, we tested Italian children in grades three and five. This age range is assumed to cover an important transition period between the use of mental and written strategies: children in third grade have yet to fully master multi-digit calculation, but by fifth grade they will have started to automatize more efficient calculation skills (Baroody and Dowker, 2003). In particular, Italian curriculum for teaching arithmetic is based on the written standard approach that requires children in grade 1 (typically 6 years old) to consolidate their counting skills and start learning the principles of adding and subtracting (*left-to-right* strategies). In second grade, the procedures for solving written additions and subtractions calculation are taught (in that order) using a columnar (*right-to-left*) strategy (Cornoldi and Lucangeli, 2004). Thus, third-grade children are expected to be reasonably skilled at single-digit computations and are presumably more likely to use simpler but less efficient strategies (e.g., counting) compared to fifth-grade children. In contrast, fifth-grade children are expected to have more efficient calculation skills—they should be more likely to accurately retrieve arithmetic facts and to have a greater knowledge of efficient strategies such as decomposition or the right-to-left algorithm (Baroody and Dowker, 2003). Thus, large differences in strategy selection could be anticipated in this age range (cf. Lemaire and Calliès, 2009; Lemaire and Brun, 2018).

Problem features also expected to influence the choice of strategy: Children are more likely to use memory retrieval for easier problems when this strategy will probably produce the right answer, whereas they choose computational strategies for more difficult problems when retrieval is less likely to generate the right answer (Siegler, 1996; Lemaire and Calliès, 2009). The likelihood of a given computational strategy being chosen thus depends on the characteristics of the problem. There is evidence to suggest that children are unlikely to use the most advanced computational strategy available to them unless the difficulty of the problem demands it. On this view, increasing the difficulty of the problem will promote the use of more advanced computational strategies because children will

maximize efficiency while preserving accuracy. Efficiency usually declines for problems that involve carrying or borrowing (Noël et al., 2001; Imbo et al., 2007), as well as for problems with increasing number of digits (Green et al., 2007; Imbo et al., 2007). Thus, to explore how problem features influenced strategy choices for these children, we manipulated the complexity of the problems (i.e., whether a borrow was required and the number of digits in the problem) and presentation format (i.e., horizontal vs. vertical).

In summary, the goals of the present experiments were to (a) compare two different methods to assess children's strategy choice; (b) replicate previous findings on how children's age effects performance in solving complex subtraction problems; and (c) test the relations among children's age, problem features, and strategy choice. Further, we used a novel method to analyze strategy choices, specifically, multinomial modeling of the frequency of strategy choices. Accordingly, our emphasis was not on strategy adaptivity (which is the focus of the choice-no-choice method); instead, we explored the development of strategy choices according children's own repertoire.

Overall, we expected to find similar patterns in both experiments. First, we predicted that fifth-graders would make more use of retrieval and less use of counting than third-graders in solving subtractions without borrowing (e.g., $24-3$). Second, we expected to see an increasingly efficient use of the decomposition strategy by older children, and a more consistent use of the right-to-left strategy, especially in problems presented in columns with borrowing. Third, we expected that children would perform better while solving single-digit no-borrow problems (e.g., $45-2$) than while solving double-digit borrow problems (e.g., $45-19$) and would use decompositions (left-to-right procedures) and standard algorithms (right-to-left procedures) more on complex problems than on simpler problems. Finally, we predicted that children would be more likely to solve horizontally-presented problems with decomposition strategies and vertically presented problems with the right-to-left algorithm (Trbovich and LeFevre, 2003; DeStefano and LeFevre, 2004).

EXPERIMENT 1

In this study, we tested children's strategy selection on multi-digit subtraction problems by means of immediate retrospective self-report. In addition to replicating the results of previous studies on complex subtraction (Torbeys and Verschaffel, 2016; Lemaire and Brun, 2018), we wanted to (a) determine the full repertoire of strategies used by children in solving this type of problem, and (b) analyze the mediating roles of children's age and problem features on strategy choice.

METHODS

Participants

Participants included 155 children: 76 in third grade (50 boys, 26 girls) with a mean age of 105.9 months ($SD = 3.8$; range = 99–112 months), and 79 in fifth-grade (42 boys, 37 girls) with a mean age of 129.8 months ($SD = 3.5$; range = 124–143 months) who

were attending Italian urban state schools. Parental consent was obtained. Children with special educational needs, intellectual disabilities, or neurological/genetic disorders, as indicated by their teachers, were not included in the study.

Materials

Arithmetical Achievement

To assess arithmetical achievement, participants were initially presented with paper-and-pencil tasks adapted from an age-standardized Italian battery (Biancardi and Nicoletti, 2004). In the *complex written calculation* test, children attempted 12 written calculation problems (4 additions, 4 subtractions, and 4 multiplications; e.g., $46+18=?$; $54-27=?$; $23\times 41=?$) without time limits. Scores were total correct (Cronbach's $\alpha = 0.78$). In the *simple calculation* test, children attempted 16 problems (8 additions and 8 subtractions) with operands between 1 and 9. For both addition and subtraction half of the results were less than 10 (e.g., $4+2=?$; $7-5=?$), and the other half were more than 10 (e.g., $10+12=?$; $30-6=?$). The total time allowed to complete the test was 200 s. Cronbach's reliability coefficients were higher than 0.80 for each set of problems.

Computer-Based Experimental Task

Children solved multi-digit subtraction problems. Two problem sets were created, each with 32 problems (see the Supplementary Material for the whole sets). In order to manipulate problem difficulty, problems were characterized by the presence/absence of borrowing procedure and by the number of digit of the subtrahend. One set required a borrow procedure in the unit position (e.g., $31-19=?$), and the other set did not require a borrow procedure (e.g., $38-26=?$). Half of each set had a subtrahend with a one-digit number (e.g., $58-6$) and the other half had a subtrahend with a two-digit number (e.g., $43-12$). The correct answers for all the 64 subtraction problems ranged from 11 to 62. Following previous literature, to control the difficulty of the individual problems, certain types were excluded (e.g. Campbell, 2005): (a) no operand had 0 or 5 as the unit digit; (b) digits were not repeated in the same decade or unit positions across operands (e.g., $64-24=?$); (c) no digits were repeated within operands (e.g., $66-31=?$); (d) no correct answers for the decades or units equaled 0 (e.g., $36-16=?$); and, finally, (e) no correct answers coincided with the second operand (e.g. $24-12=?$). Furthermore, the outcome of subtractions (i.e., odd or even numbers) and the presentation format (i.e. horizontal or vertical) were controlled. Within each set, half of the problems were assigned to the vertical presentation and half to the horizontal presentation.

Procedure

Children were tested in two sessions. At the beginning of the experimental session, in a group session lasting about 30 min, individuals' mathematical achievement was assessed with paper-and-pencil tasks adapted from the standardized Italian battery developed by Biancardi and Nicoletti (2004) in their classroom. In an individual session lasting ~60 min, the children were tested in a quiet room using the computer-based experimental task. The task was programmed using E-Prime software (Psychology

Software Tools, Inc., Pittsburgh, PA, USA) and presented on a 15-inch 1024 × 768 pixel computer screen. Problems were shown in 72-point Times New Roman black font on a white background in the center of the screen. Participants sat 60 cm from the screen. The E-Prime software controlled how long time the stimulus was displayed and recorded accuracy, response times (RTs), and the selected strategies for each trial.

Each trial began with the presentation of a fixation point (*) in the center of the computer screen for 750 ms. Then a problem was displayed (horizontally or vertically) in the center of the screen. Each trial was timed as of the moment when the problem appeared on the screen and ended when the experimenter pressed a button as promptly as possible after participants gave their answers. Problems were presented in a pseudo-random order.

Children were randomly assigned to one of two problem sets, that is, problems with or without a borrow requirement, such that there were 73 children (39 boys, 34 girls; 37 third- and 36 fifth-graders) in the no-borrow condition, and 82 (53 boys, 29 girls; 39 third-, and 43 fifth-graders) in the borrow condition.

Each participant solved 4 practice trials and 32 experimental trials. Trial-by-trial feedback on calculation accuracy was only given during the practice trials. Children were told that they would see two-digit subtraction problems (e.g., 79-37; 92-59) on the computer screen: they were asked to do the calculation aloud and to give their answers aloud, focusing equally on speed and accuracy. Immediately after having provided each solution, they were asked to verbally explain how they had reached the result (each answer was recorded).

Classification of Self-Reports

Participants' verbal self-reports were classified into five different strategy categories by two trained judges on the basis of the narrative procedure descriptions. The two judges agreed on the classification of 97% of the problems. Five main strategies emerged when children's self-reports were analyzed. Trials were categorized as: (1) *retrieval* when participants simply reported remembering or knowing the answer (); (2) *counting* when children described a sequential subtraction of a one unit at a time (e.g., 24-3 as 24, 23, 22, answer 21); (3) *left-to-right decomposition* when the answers were obtained by breaking a larger problem down into smaller ones (e.g., regrouping strategies); (4) *right-to-left algorithm* when children described arriving at the answer by first subtracting the units and then the tens (e.g., ...); (5) *other* when children reported guessing or a mixture of different procedures on the same problem.

RESULTS

Arithmetical Achievement

Performance on the two arithmetic achievement tasks was analyzed in 2 (grade) × 2 (condition: borrow, no-borrow) ANOVAs. Consistent with the use of a grade-standardized score, there were no effects of grade, and no effects of condition, indicating that children at both grades had mathematical abilities expected for their age, and that the two randomly assigned groups of children (borrow and no-borrow conditions) were

equally matched on arithmetic skills. The descriptive statistics and ANOVAs results for these analyses are presented in the Supplementary Material.

Accuracy and Response Times

Accuracy was the percentage of correct responses; response times were calculated on the basis of correct trials only.

The descriptive statistics for performance on the multi-digit subtraction task are shown in **Table 1** (upper panel). In order to verify which manipulated variables influenced the performance of multi-digit subtraction problems, response times, and accuracy were analyzed in separate 2 (complexity: one- vs. two-digit numbers in the subtrahend) by 2 (format: horizontal, vertical presentation) × 2 (grade: 3, 5), × 2 (condition: borrow, no-borrow) mixed ANOVAs, with repeated measures on the first two factors. The results of these analyses are shown in **Table 2**. For the sake of simplicity, the two-way interactions are discussed only when the three-way interactions were not significant.

As expected, the main effect of grade was significant, showing that children in third grade performed significantly worse than those in fifth grade (84 vs. 89%). There were also main effects of format, condition and complexity: children showed a better performance when problems were vertically presented (88 vs. 85%), they solved borrow problems less accurately than no-borrow problems (82 vs. 91%), and they performed less accurately on problems with double-digit subtrahends than on those with single-digit subtrahends (82 vs. 92%). These differences were confirmed also by the two-way interaction between condition and complexity: the difference between borrow and no-borrow problems was larger for problems with double-digit subtrahends (i.e., 13%; 75% vs. 88%) than for problems with single-digit subtrahends (i.e., 5%; 89% vs. 94%), although both differences were significant ($p_s < 0.001$). The significant interaction between complexity and grade indicated that the difference among grades was due to the complexity of the problems: younger children registered lower performance only when they solved subtractions with double-digit subtrahends (i.e., 78 vs. 86%; $p < 0.01$).

Finally, the interaction between complexity × format and between condition × format were significant, as well as the three-way interaction among complexity, condition, and format. These interactions revealed that the presentation format influenced children's performance only when they were asked to solve the hardest problems, that is, double-digit subtrahends involving borrowing (i.e., 71 vs. 81%; $p < 0.001$).

The analysis of response time showed significant main effects of grade, condition (borrow status), and complexity. Hence, third-graders were slower than fifth graders (17 vs. 10 s), children who solved borrow problems responded more slowly than those who solved no-borrow problems (19 vs. 9 s), and children solved problems with double-digit subtrahends more slowly than those with single-digit subtrahends (17 vs. 11 s), as highlighted by the interaction of condition × complexity ($p_s < 0.001$). The complexity × format interaction was also significant. For problems with two-digit subtrahends, children solved problems in vertical format faster than those in horizontal format (16 vs. 18 s, $p < 0.01$) whereas, for problems with single-digit

TABLE 1 | Descriptive statistics (*M* = mean; *SD* = standard deviations) of multi-digit subtraction problems.

Experiment 1									
Group	Complexity	Third grade				Fifth grade			
		Horizontal		Vertical		Horizontal		Vertical	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
ACCURACY									
No-borrow	One-digit	92.91	10.84	92.22	12.62	95.83	6.68	95.13	9.57
	Two-digit	82.43	18.97	81.08	18.31	93.40	8.70	96.53	6.42
Borrow	One-digit	89.10	13.51	87.50	13.14	90.41	13.59	89.83	13.15
	Two-digit	69.55	27.48	79.16	20.75	71.80	23.64	82.56	16.62
RESPONSE TIME									
No-borrow	One-digit	8.73	4.03	8.96	4.17	3.66	1.07	4.03	1.34
	Two-digit	15.92	7.32	14.21	4.38	6.84	2.23	6.80	2.13
Borrow	One-digit	18.01	12.95	20.04	11.41	11.45	6.52	13.07	5.98
	Two-digit	26.34	15.51	23.92	12.57	21.21	16.31	17.93	7.99
Experiment 2									
ACCURACY									
No-borrow	One-digit	92.39	16.77	90.22	16.23	94.05	11.79	94.64	10.01
	Two-digit	79.62	23.33	77.17	22.25	85.12	20.52	84.23	16.35
Borrow	One-digit	72.87	27.23	74.39	26.21	83.42	21.25	81.79	22.00
	Two-digit	46.65	32.12	52.74	32.66	65.22	31.61	68.48	24.96
RESPONSE TIME									
No-borrow	One-digit	6.82	3.51	7.60	4.70	5.28	4.71	5.46	4.72
	Two-digit	14.15	5.65	13.38	4.48	10.65	7.92	9.89	6.11
Borrow	One-digit	15.82	7.51	17.41	7.93	11.73	6.28	12.70	6.33
	Two-digit	25.17	9.93	22.60	9.20	18.79	10.30	16.41	7.05

Accuracy refers to the percentage of correct problems; response times are expressed in seconds.

TABLE 2 | Results of the mixed-design $2 \times 2 \times 2 \times 2$ ANOVAs for the accuracy and RTs, with grade (third and fifth grade) and condition (absence or presence of borrowing procedure) as the between-participants factors, and complexity (single or double-digit subtrahend) and format (horizontal or vertical presentation) as repeated measures (Experiment 1).

	Accuracy				Reaction times		
	<i>df</i>	<i>F</i>	<i>p</i>	<i>LogBF</i>	<i>F</i>	<i>p</i>	<i>LogBF</i>
Grade (G)	1,151	8.44**	0.004	13.12	28.83**	< 0.0001	40.43
Condition (Cond)	1,151	23.75**	< 0.0001	34.84	74.93**	< 0.0001	93.18
Complexity (C)	1,151	55.32**	< 0.0001	73.45	125.88**	< 0.0001	140.15
Format (F)	1,151	6.03*	0.015	9.16	1.14	0.288	1.74
G*Cond	1,151	2.58	0.110	4.18	0.06	0.802	0.10
G*C	1,151	4.84*	0.029	7.45	1.10	0.296	1.68
G*F	1,151	0.76	0.383	1.17	0.03	0.869	0.04
Cond*C	1,151	9.16**	0.003	13.90	4.60*	0.034	6.95
Cond*F	1,151	5.52*	0.020	8.52	0.09	0.765	0.14
C*F	1,151	12.75**	< 0.0001	18.95	11.62**	0.001	17.15
G*Cond*C	1,151	3.26	0.073	5.06	5.11	0.025	7.70
G*Cond*F	1,151	0.092	0.762	0.14	1.04	0.310	1.59
G*C*F	1,151	0.41	0.522	0.61	0.11	0.746	0.16
Cond*C*F	1,151	7.27*	0.008	11.20	4.13*	0.044	6.25
G*Cond*C*F	1,151	0.37	0.544	0.55	0.34	0.562	0.52

LogBF, approximated bayes factor; G, Grade; Cond, Condition; C, Complexity; F, presentation Format. * $p < 0.05$; ** $p < 0.01$.

subtrahends, they solved problems in vertical format more slowly than those in horizontal format (12 vs. 10 s, $p < 0.05$). Finally, the three-way interaction among condition, complexity, and format is shown in **Figure 1** (upper panel). Children who solved no-borrow problems did not show any effects of format, whereas those who solved borrow problems showed the interaction of complexity and format ($p_s < 0.01$).

Self-Report

The analyses of accuracy and response times showed that problem features influenced children's performance, and thus their strategy efficiency. Descriptive data on strategy choices for children in third- and fifth-grade are presented in **Table 3** which shows the number of children who used the strategy at least once and the observed frequencies across grades, conditions, complexity, and presentation format.

The Table shows the number of children who reported using each strategy at least once: the use of strategies varied significantly across ages, in particular for the more complex strategies: more older children reported left-to-right strategy than younger children (61 vs. 33%), $\chi^2(1, N = 155) = 12.07$, $p < 0.001$, Cramer's $\phi = 0.279$ whereas younger children reported to use more often the right-to-left-algorithm (93 vs. 93%), $\chi^2(1, N = 155) = 8.05$, $p = 0.005$, Cramer's $\phi = -0.228$. For the simpler strategies, the differences are less evident, but the emerged pattern seems to indicate that older children are more likely to report retrieval and less likely to use counting compared to the younger children. Thus, the overall comparison of strategy repertoire across grade shows changes as a function of children's expertise: these shifts in strategy repertoire with grade are consistent with increased access to stored arithmetic facts and a greater conceptual understanding of number.

Next, we explore the patterns of strategy selection in relation to problem features. We analyzed strategy choices in order to determine whether they varied with the same problem features as did strategy efficiency. Of interest was the frequency of strategy choice across all problems, regardless of whether those choices resulted in accurate performance. As previously mentioned, the 155 children each solved 32 subtraction problems, hence there were a total of 4,960 trials for analysis.

Analyses of strategy choices were performed using the statistical software R (R Core Team, 2015) using the following packages: *vglm* for the Multinomial models (Yee and Wild, 1996; Yee, 2015) and *Bayes Factor* for Bayesian estimates (Morey and Rouder, 2015). To determine the best fitting model for describing the relation between problem features and strategy choices, we analyzed the data with a series of multinomial models. Each model included the four independent variables: grade, condition, complexity, and presentation format. This type of discrete-choice model permits the set of choice (the four strategy options) to vary by participants and can incorporate explanatory variables that can characterize the pattern of frequencies, in this situation, strategy choice.

A model-selection strategy was performed using a procedure to detect the best-fitting model (for an example, see Fox, 2015). The type of strategy that was selected on each trial

was the dependent variable and there were four predictors: school grade attended (grade, with two levels: third and fifth grade); presence or absence of borrowing procedure (condition, with two levels: with or without borrowing); complexity of the subtrahend (subtrahend with one- or two-digits) and stimulus presentation format (two levels: horizontal or vertical). Then, starting from the null model (M0—i.e., the model including only intercepts and no predictors), we built the various models developed from all the possible combinations of the four predictors. After the null model (M0), we first explored the additive model; next all the possible two-way interactions were tested. Afterward, all the three-way interactions were explored. In total there were 14 models resulting from all the possible combination of the predictors—the saturated model with all predictors did not converge and so it was not included.

We used the likelihood ratio test to compare models, taking into consideration the Bayesian information criterion (BIC; Schwarz, 1978). In **Table 3**, the results of model comparisons are reported. ΔBIC indicates the differences between the null model (M0) and each subsequent model; a positive ΔBIC value implies that a given model is better than the null model. In order to compare the relative evidence for each different model we calculated the *Log Bayes Factor* (BF) approximations (see **Table 4**), using the formula ($\Delta BIC/2$; Raftery, 1995). For example, a *Log BF* value of 3 indicates that one model is twenty [$\exp(3) = 20$] times more likely than the null model, a difference that has been characterized as strong (Wagenmakers, 2007; Wetzels et al., 2011). More generally, the higher the ΔBIC and BF, the more likely the model is in comparison to the null model and thus provides a good fit to the data. Details of the multinomial process and the indexes that guided the model selection are given in **Table 4**.

In the first step, which involved considering additive effects only (comparable to a main effects model), including all four predictors, a positive ΔBIC value of 3,364 was found, indicating that it was a significantly better fit than the null model. This finding indicates that all four predictors influenced strategy selection. Subsequently, inclusions of two- and three-way interactions improved the overall model fit. Following this procedure, the best-fitting model was *M11* (see **Table 3**), which included the interaction of three factors, that is borrow $\times$ grade $\times$ complexity and an additive effect of presentation format. Comparing the ΔBIC values, we found that *M11* explained the data more than a million times ($\text{Log BF} = 14$) better than any of the other models. The interactive portion of the *M11* model is represented in **Figure 2** (upper panel), which shows the estimated probability for each strategy as a function of each combination of grade, complexity, and condition (borrow vs. no-borrow). The three-way interaction reflects the influence of the older children's greater experience and reveals clear differences in strategy choice across the problem features. In particular, the strategy used most frequently was the right-to-left procedure (i.e., St. Alg. In **Figure 2**): It was used on more problems than other strategies by both third- and fifth-graders on two-digit problems, but the younger children tended to

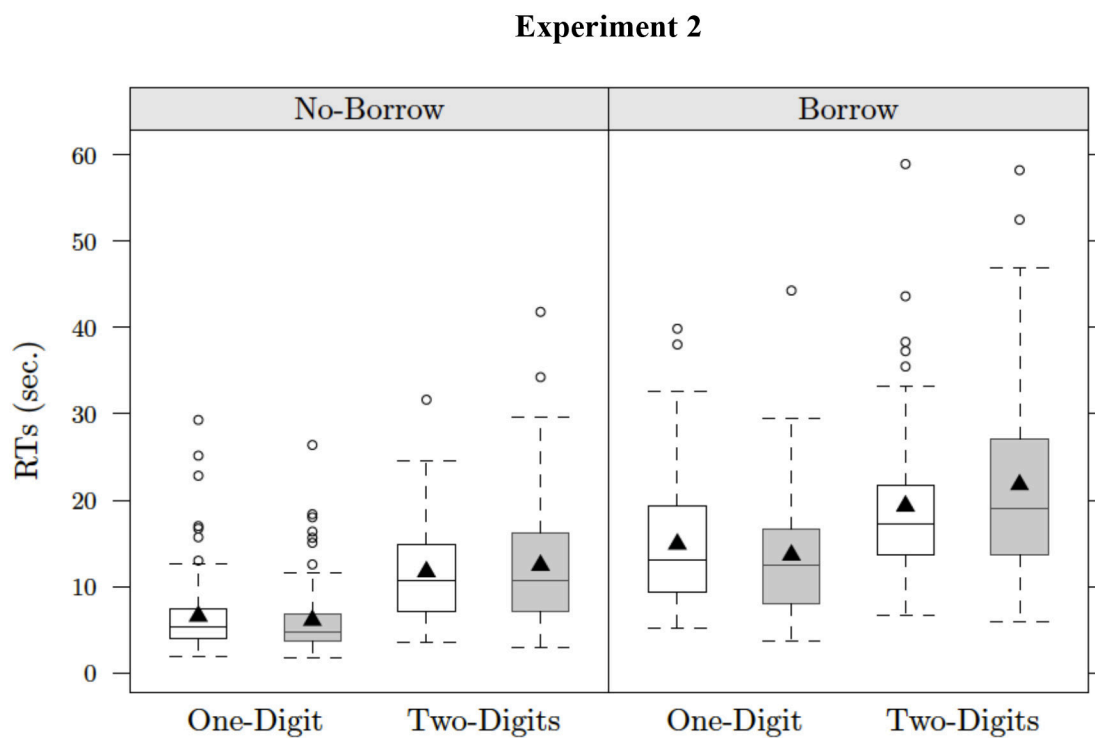
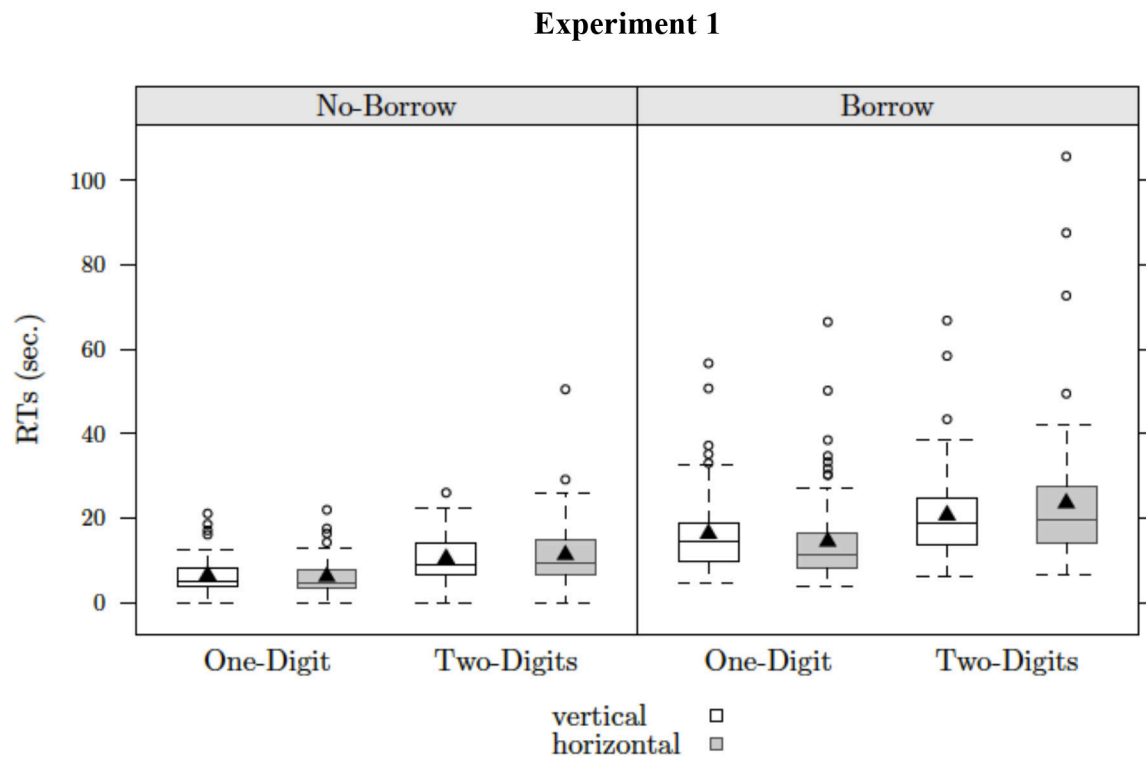


FIGURE 1 | Representation of the three-way interaction of borrow*complexity*presentation format on correct RTs for Experiment 1 and 2.

TABLE 3 | Descriptive data on strategy choices for children in third- and fifth-grade; showing the number of children who used the strategy at least once (*n*), the range, and the observed frequencies according grade, condition, complexity, and presentation format.

												Third graders										Fifth graders													
												No borrow								Borrow				No borrow								Borrow			
												Simple				Complex				Simple		Complex		Simple				Complex				Simple		Complex	
												n	Range	Col	Row	Col	Row	Col	Row	Col	Row	Col	Row	Col	Row	Col	Row	Col	Row	Col	Row	Col	Row		
Experiment 1	(n = 76)											(n = 79)																							
Retrieval	29	0–14	90	94	0	1	1	1	1	0	38	0–18	188	192	5	7	2	0	2	1															
Counting	31	0–23	47	50	2	3	28	29	10	17	28	0–16	27	28	5	0	18	24	1	2															
Left-to-right	25	0–32	6	4	6	6	60	83	40	38	48	0–32	30	31	42	52	165	182	117	137															
Right-to-Left	71	0–32	138	143	282	6	147	121	217	38	61	0–32	35	34	230	52	86	78	177	137															
Other	34	0–32	15	5	6	7	76	78	44	52	34	0–30	8	3	6	3	73	60	47	45															
Experiment 2	(n = 88)											(n = 87)																							
Retrieval	39	0-19	59	74	27	20	28	29	12	14	70	0–27	171	181	48	40	60	86	24	25															
Counting	76	0-32	121	127	98	95	115	109	59	56	59	0–30	40	44	43	35	81	77	71	52															
Left-to-right	42	0-25	28	22	30	27	61	53	64	60	59	0–26	13	28	49	65	79	77	101	129															
Right-to-Left	86	0-32	168	153	221	234	124	137	193	198	78	0–32	104	75	188	188	148	128	172	162															

n is the number of individuals in each grade who reported using the procedure at least once. Col, columnar (vertical) presentation; Row, horizontal presentation.

TABLE 4 | Model comparison for strategy choice in Experiment 1.

	<i>BIC</i>	Δ_{BIC}	<i>Log BF</i>	Model
M0	3832	0		1
M1	468	3364	37.31	pres.form + condition + grade + complexity
M2	445	3387	25.68	complexity + condition * grade + pres.form
M3	460	3371	33.40	pres.form + condition + grade * complexity
M4	478	3353	42.44	grade * pres.form + condition + complexity
M5	422	3410	14.31	pres.form + grade + complexity * condition
M6	470	3362	38.03	grade + condition * pres.form + complexity
M7	478	3354	42.29	condition + grade + pres.form * complexity
M8	432	3399	19.45	grade * pres.form + condition * complexity
M9	462	3370	34.13	condition * pres.form + complexity * grade
M10	455	3377	30.67	complexity * pres.form + grade * condition
M11	393	3438		condition * grade * complexity + pres.form
M12	466	3366	36.25	complexity + condition * pres.form * grade
M13	486	3346	46.20	condition + pres.form * grade * complexity
M14	439	3393	22.82	grade + condition * complexity * pres.form

BIC, Bayesian Information Criterion; Δ_{BIC} , BIC difference with respect to the null model (M0); *LogBF*, approximated bayes factor respect to the best model (M1) calculated through the relative likelihood [i.e., $(\Delta_{BIC}/2)$]. The higher the Δ_{BIC} the better the model. Strategy, type of strategy (i.e., counting, retrieval, decomposition, and written calculation); Grade, 3rd and 5th grade; Complexity, single or double-digit subtrahend; Condition (absence or presence of borrowing procedure); Pres.form, vertical and horizontal presentation format.

use this procedural strategy even more often than the older children.

Other differences in strategy choice were found that were also related to grade. For example, retrieval was reported more frequently by fifth- than by third-graders on the simpler problems, that is, on no-borrow problems with a single-digit subtrahend (e.g., 57– 6). Both counting and decomposition

strategies were generally reported less frequently than the right-to-left algorithm, except on one-digit borrow problems for fifth graders, where this strategy was the most frequent. Counting was reported somewhat more often by third- than by fifth-grade children, for all problems except the two-digit no-borrow problems. The left-to-right decomposition strategy was reported more often by fifth- than by third-grade children, specifically on problems with two-digit subtrahends, although the differences were modest. Finally, children's strategy reports were more likely to include a mixture of strategies (i.e., Other in **Figure 2**) borrow problems, especially one-digit ones. In summary, across grades, children showed a pattern of shifting from counting to retrieval strategies on the easier problems, that is, those with one-digit subtrahends, and a similar, but smaller shift toward left-to-right strategies on the harder problems, especially in the borrow condition. As **Figure 2** (upper panel) highlights, the presence or absence of borrowing as well as the complexity of the subtrahend interacted to determine which strategy children selected on specific problems, suggesting that they were influenced by these factors as they chose which strategies to implement.

An additive effect of presentation format was observed, showing that this feature did not interact with problem characteristics in influencing children's strategy choices: the difference in the frequency use according the vertical or horizontal presentation format was small but consistent across other combinations of predictors. This is an interesting and novel finding, because it indicates that strategy choice can be influenced both by factors inherent to the solution process (i.e., problem complexity) and by features of the visual display (i.e., format).

EXPERIMENT 2

In Experiment 2 we explored patterns of strategies chosen by children using an extended forced-choice condition. As

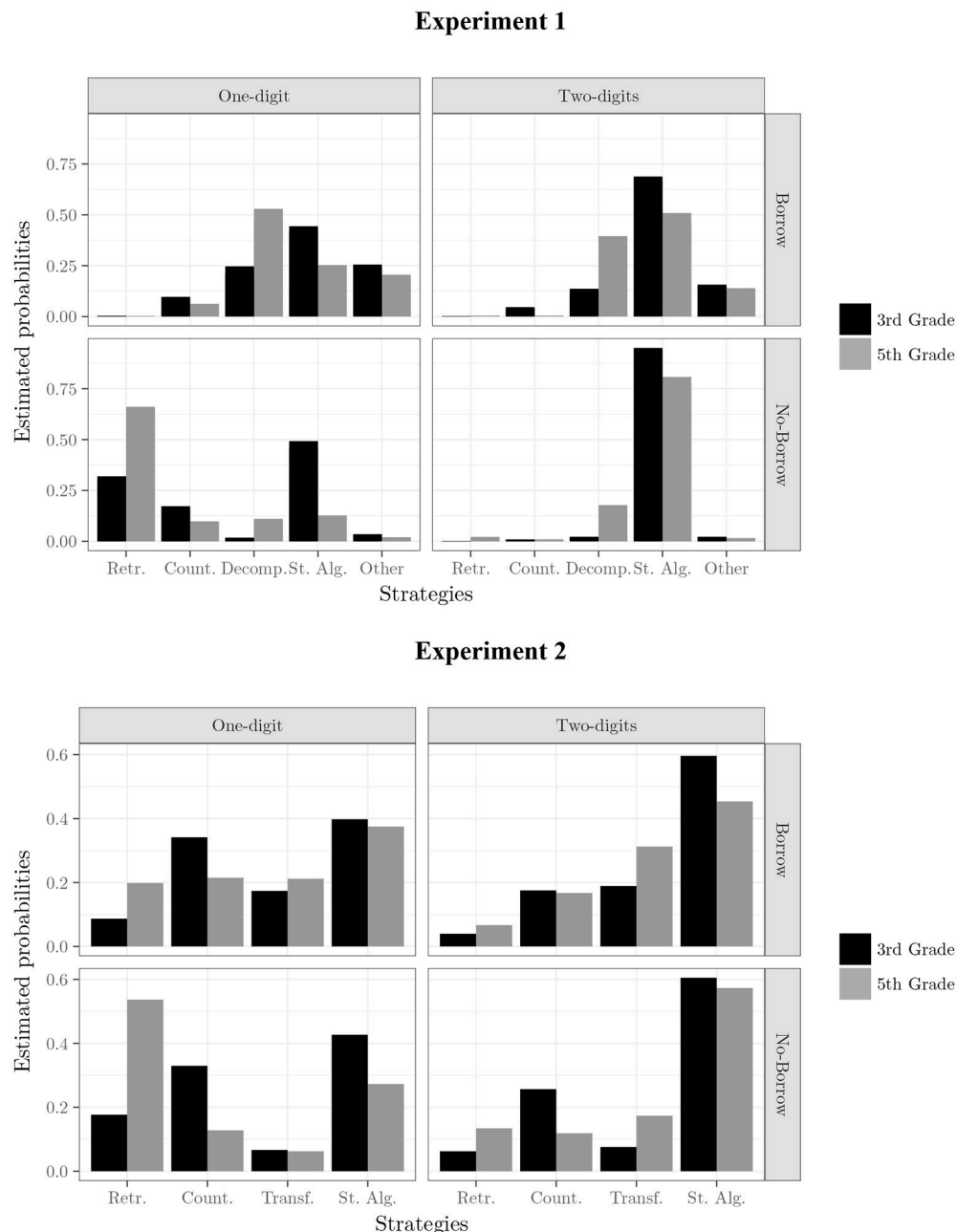


FIGURE 2 | Principal effects of the best model for prediction of strategy choice in both Experiments. Figures show a representation of the estimated probability of strategy choice according the interaction of grade*borrow*complexity. Presentation format is not reported in the figure because it enters in the models as additive factor. Retr., retrieval; Count., counting; Decomp., left-to-right procedure; St. Alg., right-to-left procedure; Other, confusing reporting.

in Experiment 1, children performed the same two tasks—the paper-and-pencil tasks assessing arithmetical achievement and the computerized mental subtraction task—with the only difference related to the collection of the strategy used. In this experiment, children were asked to choose the strategies used among a repertoire of alternatives based on the solution procedures resulted from Experiment 1. The goals of this experiment were to (a) replicate the results emerged in Experiment 1 and (b) determine whether the pattern of factors

that emerged in the multinomial analysis is generalizable to data collected via another method of assessing strategy choice.

METHODS

Participants

Participants included 175 children: 88 in third grade (47 boys, 41 girls) with a mean age of 100.2 months ($SD = 3.6$; range = 93–107.5 months), and 87 in fifth-grade (39 boys, 48 girls) with

a mean age of 124.7 months ($SD = 4.3$; range = 109–140.5 months) who were attending Italian urban state schools. Parental consent was obtained. Children with special educational needs, specific learning disorders, intellectual disabilities, or neurological/genetic disorders, as indicated by their teachers, were not included in the study.

Procedure

The design and procedure exactly replicated those of Experiment 1, with the sole exception being the way children strategy choices were collected. Children were asked to indicate how they had solved each problem by choosing one of four strategies (counting, retrieval, left-to-right decomposition, or right-to-left algorithm), which were explained with examples at the beginning of the individual session. Thus, in the present research, after completing each operation, participants were asked to indicate which out of four strategies they had used to solve each problem.

Children were randomly assigned to either the no-borrow or borrow condition, such that there were 89 children (45 boys, 44 girls; 48 third- and 41 fifth-graders) in the no-borrow condition, and 86 (41 boys, 45 girls; 40 third-, and 46 fifth-graders) in the borrow condition.

RESULTS

Arithmetical Academic Achievement

Performance on the arithmetic achievement tasks was analyzed in 2 (grade) $\times$ 2 (condition: borrow, no-borrow) ANOVAs. No differences were observed neither in relation to the grade nor to the assigned condition (see Supplementary Material).

Accuracy and Response Times

As in Experiment 1, the descriptive statistics for performance on the multi-digit subtraction task are shown in **Table 1** (lower panel). Both percentage of correct responses and correct mean latency were examined with separate 2 (complexity: one- vs. two-digit numbers in the subtrahend) by 2 (format: horizontal, vertical presentation) $\times$ 2 (grade: 3, 5), $\times$ 2 (condition: borrow, no-borrow) mixed ANOVAs, with repeated measures on the first two factors. The results of these analyses are shown in **Table 5**.

Regarding the accuracy, the main effect of grade was significant, showing that younger children performed significantly worse than the older ones (73 vs. 82%). Consistent with Experiment 1, the effects of condition and complexity were significant: No-borrow problems were easier to solve than borrow problems (87 vs. 68%), and problems with one-digit subtrahends were easier to solve than those with two-digit subtrahends (85 vs. 70%). In contrast to Experiment 1, there were no significant effects of format. The only significant interaction was condition $\times$ complexity which confirmed that difference between borrow and no-borrow problems was larger for more complex problems (i.e., two-digit: +24%; 58 vs. 82%; one-digit: +15%; 78 vs. 93%; $p_s < 0.001$).

Similar to the results for accuracy, there were significant main effects of grade, condition (borrow status), and complexity in the analysis of response time. Hence, third-graders were slower than fifth graders (15 vs. 11 s), children who solved borrow

problems responded more slowly than those who solved no-borrow problems (17 vs. 9 s), and children solved problems with double-digit subtrahends more slowly than those with single-digit subtrahends (16 vs. 10 s). The interaction of grade $\times$ complexity was significant. At both age groups, children were faster to solve problems with one-digit than those with two-digits subtrahends ($p_s < 0.001$), however this difference was larger for children in third grade than those in fifth grade (12 vs. 19 s for third graders and 9 vs. 14 s for fifth graders). As in Experiment 1, the complexity $\times$ format and the condition $\times$ complexity $\times$ format interactions were significant. In particular, children did not show any effects of format during the execution of no-borrow problems whereas when children solved borrow problems, they were faster in vertical than in horizontal format with two-digit problems (15 vs. 17 s, $p < 0.001$) and slower in vertical format than in horizontal format with single-digit problems (11 vs. 9 s, $p < 0.001$). The three-way interaction among condition, complexity and format is shown in **Figure 1** (lower panel).

Strategy Choice

As in the previous experiment, we analyzed strategy choices in order to determine whether they varied with the same problem features as did strategy efficiency, regardless of whether those choices resulted in accurate performance (see Supplementary Materials for the observed frequency of the strategies). The descriptive data on strategy choices are reported in **Table 3**. As for Experiment 1, the number of children who reported using each strategy at least once varied significantly across ages for all four strategies. More older children reported retrieval than younger children (80 vs. 44%), χ^2 (1, $N = 175$) = 24.38, $p < 0.001$, Cramer's phi = -0.373 whereas fewer older children reported counting than younger children (68 vs. 86%), χ^2 (1, $N = 175$) = 8.53, $p = 0.003$, Cramer's phi = 0.221. For the more complex strategies, more older than younger children reported using the left-to-right strategy (68 vs. 48%), χ^2 (1, $N = 175$) = 7.23, $p < 0.007$, Cramer's phi = -0.203. Finally, although a majority of children in both grades reported using the right-to-left algorithm, more younger than older children used the strategy at least once (98 vs. 90%), χ^2 (1, $N = 175$) = 4.84, $p = 0.023$, Cramer's phi = 0.166. Thus, in line with the previous experiment, the overall comparison of strategy repertoire across grade shows changes as a function of children's expertise. In the next paragraph, we analyze the patterns of strategy selection according to problem features.

Multinomial models and a model-selection strategy were used to analyze strategy choices in relation to problem features and grade on a trial-by-trial basis ($175 \times 32 = 5600$ trials), as described in Experiment 1. Details of the multinomial process and the indexes that guided the model selection are reported in **Table 6**.

These analyses precisely confirmed the former results: the best-fitting model was *M11*, which included the interaction of the three same factors, borrow $\times$ grade $\times$ complexity, and an additive effect of presentation format. Comparing the ΔBIC values, we found that *M11* explained the data more than 2,900 times ($\log BF = 8$) better than any of the other models. **Figure 2** (lower panel) shows the estimated probability of this model

TABLE 5 | Results of the mixed-design $2 \times 2 \times 2 \times 2$ ANOVAs for the accuracy and RTs, with grade (third and fifth grade) and condition (absence or presence of borrowing procedure) as the between-participants factors, and complexity (single or double-digit subtrahend) and format (horizontal or vertical presentation) as repeated measures (Experiment 2).

	df	Accuracy			Reaction times		
		F	P	LogBF	F	P	LogBF
Grade (G)	1,171	9.95**	0.002	14.75	17.86**	< 0.0001	25.98
Condition (Cond)	1,171	45.64**	< 0.0001	61.95	78.99**	< 0.0001	99.52
Complexity (C)	1,171	104.24**	< 0.0001	123.88	397.57**	< 0.0001	315.29
Format (F)	1,171	0.27	0.604	0.31	3.47	0.064	5.25
G*Cond	1,171	2.24	0.136	3.41	1.99	0.160	3.04
G*C	1,171	3.50	0.063	5.29	8.63**	0.004	12.91
G*F	1,171	0.04	0.841	11.77	0.41	0.521	1.53
Cond*C	1,171	7.87**	0.006	0.06	1.01	0.318	0.64
Cond*F	1,171	2.89	0.091	4.36	1.36	0.245	2.08
C*PF	1,171	0.99	0.322	1.52	26.18**	< 0.0001	37.15
G*Cond*C	1,171	0.66	0.419	1.00	0.03	0.854	0.05
G*Cond*F	1,171	1.53	0.217	2.34	0.01	0.921	0.02
G*C*F	1,171	0.01	0.908	0.02	0.53	0.466	0.82
Cond*C*F	1,171	2.10	0.150	3.20	6.62*	0.011	9.97
G*Cond*C*F	1,171	0.04	0.843	0.06	0.01	0.916	0.02

Strategy, type of strategy (i.e., counting, retrieval, decomposition, and written calculation); Grade: 3rd and 5th grade; Complexity, single or double-digit subtrahend; Condition (absence or presence of borrowing procedure); Pres.form, vertical and horizontal presentation format.

for each strategy as a function of each combination of grade, complexity, and condition (borrow vs. no-borrow). The overall pattern confirmed the strategy used most frequently was the right-to-left procedure: It was used on more problems than other strategies by both third-graders on all four types of problems and by fifth-graders on all except single-digit no-borrow problems, which were very frequently solved with retrieval. Retrieval was reported more frequently by fifth- than by third-graders on all problems, especially the simpler ones, and both counting and decomposition strategies were generally reported less frequently than the right-to-left algorithm. Compared to Experiment 1, counting was generally used more often, especially by third-grade children, for all problems except the hardest (i.e., two-digit borrow problems). The left-to-right decomposition strategy was reported somewhat more often by fifth- than by third-grade children, specifically on problems with two-digit subtrahends, although it was less used compared to Experiment 1. These analyses confirmed that the presence or absence of borrowing as well as the complexity of the subtrahend interacted with children's expertise to determine which strategy they selected on a specific problem.

This pattern of results strengthens the secondary role of presentation format that seems not to directly influence children's strategy choices: small differences in the frequency use emerged according the vertical or horizontal format and, above all, consistent across other combinations of predictors.

DISCUSSION

Children use a variety of strategies to solve mathematical problems (e.g., Barrouillet et al., 2008). Their strategy repertoire

is assumed to reflect an integrated network of conceptual and procedural knowledge that allows them to decide how to perform a strategy, when to use it, and why (Hiebert and Lefevre, 1986; Bisanz and LeFevre, 1990). The goal of the present research was to explore key factors that influence children's strategy choices on multi-digit subtraction problems and to directly compare two different methods for assessing children's strategy choices. To achieve this end, two different experiments were conducted on similar cohorts of third- and fifth-grade children: In the first experiment, strategy selection was investigated by means of a *free-choice* (verbal self-report) paradigm, whereas in the second study a *discrete-choice* approach was implemented. Problem features, such as complexity (i.e., whether there were one- or two-digit subtrahends) and whether the solution crossed a decade boundary (i.e., required a borrow operation) were manipulated, in addition to presentation format (i.e., horizontal vs. vertical alignment). Classical statistical analyses were applied to children's performance (i.e., accuracy and response times), and multinomial models were used to analyze strategy choices in relation to problem features and grade on a trial-by-trial basis.

Analyses of accuracy and response times in both experiments showed typical age-related improvement in performance: Fifth-grade children solved problems more quickly and accurately than third-grade children. Children's accuracy was sensitive to problem features that influence the difficulty of the problem, specifically, children assigned to the borrow condition correctly solved fewer problems than children assigned to the no-borrow condition and both groups were less accurate in solving problems with a double-digit subtrahend. A comparison of the results of the two studies revealed a discrepancy related to the contextual feature: In the first experiment children's performance was

TABLE 6 | Model comparison for strategy choice in Experiment 2.

	<i>BIC</i>	Δ_{BIC}	<i>Log BF</i>	<i>Model</i>
M0	1453	0		1
M1	401	1052	16.20	pres.form + condition + grade + complexity
M2	385	1068	8.32	complexity + condition * grade + pres.form
M3	391	1061	11.55	pres.form + condition + grade * complexity
M4	400	1053	15.89	grade * pres.form + condition + complexity
M5	398	1055	14.59	pres.form + grade + complexity * condition
M6	406	1047	18.75	grade + condition * pres.form + complexity
M7	402	1051	16.66	condition + grade + pres.form * complexity
M8	397	1056	14.27	grade * pres.form + condition * complexity
M9	397	1056	14.09	condition * pres.form + complexity * grade
M10	386	1067	8.78	complexity * pres.form + grade * condition
M11	368	1084		condition * grade * complexity + pres.form
M12	396	1057	13.70	complexity + condition * pres.form * grade
M13	399	1054	15.29	condition + pres.form * grade * complexity
M14	409	1043	20.56	grade + condition * complexity * pres.form

BIC, Bayesian Information Criterion; Δ_{BIC} , *BIC* difference with respect to the null model (M0); *LogBF*, approximated bayes factor respect to the best model (M11) calculated through the relative likelihood [i.e., ($\Delta_{BIC}/2$)]. The higher the Δ_{BIC} the better the model. *Strategy*, type of strategy (i.e., counting, retrieval, decomposition, and written calculation); *Grade*, 3rd and 5th grade; *Complexity*: single or double-digit subtrahend; *Condition* (absence or presence of borrowing procedure); *Pres.form*, vertical and horizontal presentation format.

influenced by the presentation format only when they had to solve the hardest problems (vertical presentation improved correct responses); whereas, in the second experiment, children's accuracy was not sensitive to presentation format. Children's latencies, in contrast, were related to all of the problem features, and showed the same pattern of significant effects in both experiments. Increased complexity (both in terms of borrowing procedure and subtrahend size) slowed problem execution.

Presentation format also influenced solution latencies in relation to problem difficulty: Children were faster to correctly solve double-digit borrow problems presented in columns than in rows, whereas the reverse pattern was found for single-digit problems. The differential efficiency of performance shown on correct latencies (i.e., for borrow problems, children were faster in horizontal format for one-digit problems such as 73–5 but faster with vertical format for two-digit problems such as 43–29) suggests that choices were not strategic, per se, but were driven more directly by problem format. This conclusion is consistent with the absence of any interactions between presentation format and either grade or complexity on strategy choice. Other research has suggested that different working memory resources may be implicated as a function of presentation format (e.g., Trbovich and LeFevre, 2003; Caviola et al., 2012). Thus, manipulation of presentation format may influence strategy choices independently of factors that are related to expertise or problem complexity.

The increased level of performance with age corresponds to similar patterns found in previous research (see Campbell, 2005; Cohen Kadosh and Dowker, 2015 for a general overviews), such that children's performance on complex subtraction problems

is linked to their level of experience (i.e., school grade) and to variability in problem features that reflect computational processes (Imbo and Vandierendonck, 2007c; Lemaire and Calliès, 2009). Novel results were obtained for presentation format where effects occurred only on borrow problems and varied with complexity. These patterns were further qualified by the analyses of strategy choice, as described below. Further research on the relations between superficial features and those tied directly to computational demands may have important implications for understanding children's solution processes on complex problems. For example, it would be interesting to better understand how different combinations of characteristics, such as problems presented in other familiar formats (e.g., auditory; Noël et al., 1997; LeFevre et al., 2001), may also influence accuracy and response times.

The second novel and interesting set of results concerns children's strategy choice. No previous research defined in detail the full range of strategies used by children to solve complex subtraction problems. In both experiments, we analyzed strategy choices using multinomial modeling in which all factors, that is, expertise (i.e., grade), complexity of problem, condition (i.e., borrow vs. no-borrow) and presentation format were included as predictors. Interestingly, the best-fitting model was the same in both experiments and included a three-way interaction of grade, condition, and complexity, and an additive effect of presentation format. First, consider problems with one-digit subtrahends. We observed in both studies that fifth-grade children choose retrieval more than third graders on no-borrow problems (e.g., 78–5) whereas third-grade children were more likely to choose the standard algorithm. In contrast, third-grade children chose counting more often than fifth graders on both borrow (e.g., 73–5) and no-borrow problems (e.g., 89–7). These patterns for single-digit subtractions show a shift from less- to more-sophisticated strategies with expertise (i.e., more retrieval, less counting), accompanied by a higher reliance on algorithmic solutions by the younger children.

For the more difficult problems with two-digit subtrahends, compared to fifth-graders, third graders chose counting more often on no-borrow problems (e.g., 68–41), and the standard algorithm more often on borrow problems (e.g., 43–29). Compared to third-graders, fifth graders more often chose decompositions for both borrow and no-borrow problems. Again, these patterns of strategy choice, emerged in both studies, indicate that older children, relied more on strategies that were efficient (i.e., less use of counting) and reflected their superior conceptual understanding (i.e., more use of decompositions). It is worth to remember that these differences which emerged in strategies selection may reflect a schooling or recency effect (Lemaire and Brun, 2018): third graders may be more likely to choose a standard (written) algorithm solution because it is a strategy that they recently learnt at school (it is taught during the second and third grades in Italy), whereas older children can rely on more efficient strategies (i.e., decomposition) linked to a better mastery of basic arithmetic knowledge.

At a more general level, multinomial modeling of strategy choices confirmed the importance of some key influences on children's strategy choices for subtraction. The presence

or absence of borrowing, the value of the subtrahend, and presentation format all influenced which strategy was adopted to solve the problems. Moreover, in line with our expectations, children varied in their strategy repertoires and their use of those strategies according to their level of experience and in relation to problem complexity. These findings extend results reported in previous studies, encompassing a wider range of subtraction problems. Previous research on simple addition problems indicated that children tend to shift from counting (an inefficient procedural strategy) to more efficient memory-based retrieval with increased age (Widaman et al., 1992; Lemaire and Siegler, 1995; Geary, 2004; Reed et al., 2015). The present results support a similar pattern for complex subtraction, but show that shifts are related to problem features and that there is also considerable persistence in strategy availability with development from grades three to five. Taken together, these outcomes support Siegler's overlapping waves model (1996), at least for the retrieval vs. non-retrieval strategies: Children do not simply use a particular strategy until a better one is available, instead they have many strategies at once and it is frequency of use that changes across development (Lemaire and Siegler, 1995).

In previous research, children also showed developmental and educational changes in the use of left-to-right and right-to-left strategies (Fuson et al., 1988; Geary et al., 2004). Both the methods we used to assess strategy choice showed that children have a wide repertoire of strategies that overlapped from third- to fifth-grade, and that, although strategy choice changed with grade, it also depended heavily on problem features. Thus, the present study was consistent with the findings of persistent diverse strategy use across expertise, a finding observed even among adults solving simple arithmetic problems (e.g., LeFevre et al., 1996; Barrouillet et al., 2008), and extended the conclusion that children do not use a single strategy to solve two-digit subtraction problems (Lemaire and Calliès, 2009). Thus, our findings replicated similar patterns from previous studies and extended the overlapping waves model to a wider repertoire of strategies. In fact, these four strategy categories were often used in previous studies and account well for data observed in adults (e.g., Campbell and Xue, 2001; Campbell and Austin, 2002; LeFevre et al., 2006) whereas all of them were never considered together before in a developmental sample.

Finally, the present research shows the validity of two different self-report methods for assessing strategy choice: Both *free-choice* and *discrete-choice* approaches provided valuable information about strategy repertoire and strategy choices. Another important contribution of the present work is the use of a novel analysis of categorical data on strategy choices on a problem-by-problem basis. Together, the combination of self-report method and categorical analyses of those self-reports allowed us to document developmental changes in relation to different problem features that are known to influence strategy efficiency. Previously, the use of strategies has often been examined in terms of strategy efficiency and adaptivity (Lemaire et al., 2000), which refers to the speed and accuracy with which strategies are implemented: A multilevel modeling approach extends the analyses of these aspects from an individual to an item level. Future studies may apply this approach which allows a

sufficient amount of data for establishing temporal and accuracy characteristics of the strategies in a reliable way (Luwel et al., 2009).

As always, this research had limitations. First, as noted by Lemaire and Brun (2018), allowing students to have full choice of strategies does not allow an unbiased investigation of strategy execution and efficiency and so future studies should specifically address this limitation. Second, further research should explore how individual differences in cognitive resources can differently affect the pattern of strategies that students select and apply to different types of problems, maybe also including both simple and complex problems in a within subject design. For example, researchers have shown that children with mathematical difficulties distribute working memory resources differently than do their typically-developing peers (Mammarella et al., 2013a,b). Future research should address this important issue because it has clear implications for scenarios outside the experimental setting, such as in teaching decisions (e.g., when teachers have to choose whether to focus on practice or on exploration and flexibility; Imbo and LeFevre, 2009), and in the clinical setting (e.g., for the development of effective intervention programs; Caviola et al., 2016). It is generally assumed that children experiencing mathematical learning difficulties find it difficult to use both retrieval and right-to-left strategies (see Geary, 2004, for a review). But a more in-depth knowledge of which strategies prove more efficient in relation to a problem's complexity and an individual's resources might help to improve such children's mathematical achievement and may be beneficial in the design of appropriate diagnostic tools and educational interventions. Finally, it is important to conduct cross-cultural studies to understand how cultural and schooling effects may influence strategy selection for children of various ages (Imbo and LeFevre, 2009, 2011).

In brief, the present research showed that there is great variability in strategy selection in complex subtraction problems and revealed important effects of grade, problem complexity, and presentation format on how participants solve complex arithmetic problems, both in terms of performance and in choice of strategies. We found that problem features influence performance, either because these physical qualities compromise the efficiency of strategies that they usually applied in mental calculation or because one or more of these features directly influences strategy choice.

ETHICS STATEMENT

The protocol was approved by the Ethics Committee of the University of Padova. Consent forms were sent to the parents. If the parents agreed to participate, they were sent them forms to fill in and their children were examined at their respective schools. Children generally reported the tasks within the realm of what they normally do in the school day.

AUTHOR CONTRIBUTIONS

SC and IM developed the study concept. SC organized and supervised data collection. MP conducted the statistical analyses.

SC and J-AL drafted the manuscript. IM and J-AL provided critical revisions. All authors approved the final version of the manuscript for submission.

FUNDING

This project has been partially supported by the European Union's Horizon 2020 research and innovation

programme under the Marie Skłodowska-Curie grant agreement No. 700031.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01209/full#supplementary-material>

REFERENCES

- Ardiale, E., and Lemaire, P. (2013). Effects of execution duration on within-item strategy switching in young and older adults. *J. Cogn. Psychol.* 25, 464–472. doi: 10.1080/20445911.2013.789854
- Baroody, A. J., and Dowker, A. (2003). *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise*. Mahwah, NJ: Erlbaum.
- Barrouillet, P., Mignon, M., and Thevenot, C. (2008). Strategies in subtraction problem solving in children. *J. Exp. Child Psychol.* 99, 233–251. doi: 10.1016/j.jecp.2007.12.001
- Beishuizen, M. (1993). Mental strategies and materials or models for addition and subtraction up to 100 in Dutch second grades. *J. Res. Math. Educ.* 294–323.
- Biancardi, A., and Nicoletti, C. (2004). *Batteria per la Discalculia Evolutiva (BDE)*. Torino: Omega.
- Bisanz, J., and LeFevre, J. A. (1990). "Strategic and nonstrategic processing in the development of mathematical cognition," in *Children's Strategies: Contemporary Views of Cognitive Development*, ed D. F. Bjorklund (Hillsdale, NJ: Erlbaum), 213–244.
- Campbell, J. I. D. (2005). *Handbook of Mathematical Cognition*. New York, NY: Psychology Press.
- Campbell, J. I. D., and Austin, S. (2002). Effects of response time deadlines on adults' strategy choices for simple addition. *Memory Cogn.* 30, 988–994. doi: 10.3758/BF03195782
- Campbell, J. I. D., and Xue, Q. (2001). Cognitive arithmetic across cultures. *J. Exp. Psychol.* 130, 299–315. doi: 10.1037/0096-3445.130.2.299
- Caviola, S., Gerotto, G., and Mammarella, I. C. (2016). Computer-based training for improving mental calculation in third- and fifth-graders. *Acta Psychol.* 171, 118–127. doi: 10.1016/j.actpsy.2016.10.005
- Caviola, S., Mammarella, I. C., Cornoldi, C., and Lucangeli, D. (2012). The involvement of working memory in children's exact and approximate mental addition. *J. Exp. Child Psychol.* 112, 141–160. doi: 10.1016/j.jecp.2012.02.005
- Caviola, S., Mammarella, I. C., Lucangeli, D., and Cornoldi, C. (2014). Working memory and domain-specific precursors predicting success in learning written subtraction problems. *Learn. Individ. Differ.* 36, 92–100. doi: 10.1016/j.lindif.2014.10.010
- Cohen Kadosh, R., and Dowker, A. (2015). *The Oxford Handbook of Numerical Cognition*. Oxford: Oxford University Press.
- Cornoldi, C., and Lucangeli, D. (2004). Arithmetic education and learning disabilities in Italy. *J. Learn. Disabil.* 37, 42–49. doi: 10.1177/00222194040370010501
- Davis, H., and Carr, M. (2001). Gender differences in mathematics strategy use: the influence of temperament. *Learn. Individ. Differ.* 13, 83–95. doi: 10.1016/S1041-6080(02)00063-8
- DeStefano, D., and LeFevre, J. A. (2004). The role of working memory in mental arithmetic. *Eur. J. Cogn. Psychol.* 16, 353–386. doi: 10.1080/09541440244000328
- Ericsson, K. A., and Simon, H. A. (1993). *Protocol Analysis: Verbal Reports as Data (Revised Edition)*. Cambridge, MA: MIT Press.
- Fox, J. (2015). *Applied Regression Analysis and Generalized Linear Models*. Thousand Oaks, CA: Sage Publications.
- Fuson, K. C., Stigler, J. W., and Bartsch, K. (1988). Grade placement of addition and subtraction topics in Japan, mainland China, the Soviet Union, Taiwan, and the United States. *J. Res. Math. Educ.* 19, 449–456. doi: 10.2307/749177
- Geary, D. C. (2004). Mathematics and learning disabilities. *J. Learn. Disabil.* 37, 4–15. doi: 10.1177/00222194040370010201
- Geary, D. C., Hamson, C. O., and Hoard, M. K. (2000). Numerical and arithmetical cognition: a longitudinal study of process and concept deficits in children with learning disability. *J. Exp. Child Psychol.* 77, 236–263. doi: 10.1006/jecp.2000.2561
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., and DeSoto, M. C. (2004). Strategy choice in simple and complex addition: contributions of working memory and counting knowledge for children with mathematical disability. *J. Exp. Child Psychol.* 88, 121–151. doi: 10.1016/j.jecp.2004.03.002
- Green, H. J., Lemaire, P., and Dufau, S. (2007). Eye movement correlates of younger and older adults' strategies for complex addition. *Acta Psychol.* 125, 257–278. doi: 10.1016/j.actpsy.2006.08.001
- Hiebert, J., and Lefevre, P. (1986). "Conceptual and procedural knowledge in mathematics: an introductory analysis," in *Conceptual and Procedural Knowledge: The Case of Mathematics*, ed J. Hiebert (Hillsdale, NJ: Erlbaum), 1–27.
- Imbo, I., and LeFevre, J.-A. (2009). Cultural differences in complex addition: efficient Chinese versus adaptive Belgians and Canadians. *J. Exp. Psychol. Learn. Mem. Cogn.* 35, 1465–1476. doi: 10.1037/a0017022
- Imbo, I., and LeFevre, J.-A. (2010). The role of phonological and visual working memory in complex arithmetic for Chinese- and Canadian-educated adults. *Mem. Cognit.* 38, 176–185. doi: 10.3758/MC.38.2.176
- Imbo, I., and LeFevre, J. A. (2011). Cultural differences in strategic behavior: a study in computational estimation. *J. Exp. Psychol.* 37:1294. doi: 10.1037/a0024070
- Imbo, I., and Vandierendonck, A. (2007a). The role of phonological and executive working memory resources in simple arithmetic strategies. *Eur. J. Cogn. Psychol.* 19, 910–933. doi: 10.1080/09541440601051571
- Imbo, I., and Vandierendonck, A. (2007b). The development of strategy use in elementary school children: working memory and individual differences. *J. Exp. Child Psychol.* 96, 284–309. doi: 10.1016/j.jecp.2006.09.001
- Imbo, I., and Vandierendonck, A. (2007c). Do multiplication and division strategies rely on executive and phonological working memory resources? *Mem. Cognit.* 35, 1759–1771. doi: 10.3758/BF03193508
- Imbo, I., and Vandierendonck, A. (2008). Effects of problem size, operation, and working-memory span on simple-arithmetic strategies: differences between children and adults? *Psychol. Res.* 72, 331–346. doi: 10.1007/s00426-007-0112-8
- Imbo, I., Vandierendonck, A., and Vergauwe, E. (2007). The role of working memory in carrying and borrowing. *Psychol. Res.* 71, 467–483. doi: 10.1007/s00426-006-0044-8
- Kilpatrick, J., Swafford, J., and Findell, B. (2001). *Adding It Up: Helping Children Learn Mathematics*. Washington, DC: National Academy Press.
- Kirk, E. P., and Ashcraft, M. H. (2001). Telling stories: the perils and promise of using verbal reports to study math strategies. *J. Exp. Psychol.* 27:157. doi: 10.1037/0278-7393.27.1.157
- LeFevre, J. A., DeStefano, D., Penner-Wilger, M., and Daley, K. E. (2006). Selection of procedures in mental subtraction. *Can. J. Exp. Psychol.* 60, 209–220. doi: 10.1037/cjcp2006020
- LeFevre, J. A., Lei, Q., Smith-Chant, B. L., and Mullins, D. B. (2001). Multiplication by eye and by ear for Chinese-speaking and English-speaking adults. *Can. J. Exp. Psychol.* 55, 277–284. doi: 10.1037/h0087374
- LeFevre, J. A., Sadesky, G. S., and Bisanz, J. (1996). Selection of procedures in mental addition: reassessing the problem size effect in adults. *J. Exp. Psychol.* 22, 216–230. doi: 10.1037/0278-7393.22.1.216

- Lemaire, P., and Brun, F. (2018). Age-related changes in children's strategies for solving two-digit addition problems. *J. Num. Cogn.* 3, 582–597. doi: 10.5964/jnc.v3i3.117
- Lemaire, P., and Calliès, S. (2009). Children's strategies in complex arithmetic. *J. Exp. Child Psychol.* 103, 49–65. doi: 10.1016/j.jecp.2008.09.007
- Lemaire, P., and Lecacheur, M. (2011). Age-related changes in children's executive functions and strategy selection: a study in computational estimation. *Cogn. Dev.* 26, 282–294. doi: 10.1016/j.cogdev.2011.01.002
- Lemaire, P., Lecacheur, M., and Farioli, F. (2000). Children's strategy use in computational estimation. *Can. J. Exp. Psychol.* 54, 141–148. doi: 10.1037/h0087336
- Lemaire, P., and Siegler, R. S. (1995). Four aspects of strategic change: contributions to children's learning of multiplication. *J. Exp. Psychol.* 124, 83–97. doi: 10.1037/0096-3445.124.1.83
- Linsen, S., Verschaffel, L., Reynvoet, B., and De Smedt, B. (2014). The association between children's numerical magnitude processing and mental multi-digit subtraction. *Acta Psychol.* 145, 75–83. doi: 10.1016/j.actpsy.2013.10.008
- Luwel, K., Onghena, P., Torbeyns, J., Schillemans, V., and Verschaffel, L. (2009). Strengths and weaknesses of the choice/no-choice method in research on strategy use. *Eur. Psychol.* 14, 351–362. doi: 10.1027/1016-9040.14.4.351
- Mabbott, D. J., and Bisanz, J. (2003). Developmental change and individual differences in children's multiplication. *Child Dev.* 74, 1091–1107. doi: 10.1111/1467-8624.00594
- Mammarella, I. C., Bomba, M., Caviola, S., Broggi, F., Neri, F., Lucangeli, D., et al. (2013a). Mathematical difficulties in nonverbal learning disability or co-morbid dyscalculia and dyslexia. *Dev. Neuropsychol.* 38, 418–432. doi: 10.1080/87565641.2013.817583
- Mammarella, I. C., Caviola, S., Cornoldi, C., and Lucangeli, D. (2013b). Mental additions and verbal-domain interference in children with developmental dyscalculia. *Res. Dev. Disabil.* 34, 2845–2855. doi: 10.1016/j.ridd.2013.05.044
- Morey, R. D., and Rouder, J. N. (2015). *BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-2*. Available online at: <https://CRAN.R-project.org/package=BayesFactor>
- Noël, M.-P., Désert, M., Aubrun, A., and Seron, X. (2001). Involvement of short-term memory in complex mental calculation. *Mem. Cognit.* 29, 34–42. doi: 10.3758/BF03195738
- Noël, M.-P., Fias, W., and Brysbaert, M. (1997). About the influence of the presentation format on arithmetical-fact retrieval processes. *Cognition* 63, 335–374. doi: 10.1016/S0010-0277(97)00009-7
- Noël, M.-P., Seron, X., and Trovarelli, F. (2004). Working memory as a predictor of addition skills and addition strategies in children. *Curr. Psychol. Cogn.* 22, 3–25.
- R Core Team (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available online at: <https://www.R-project.org/>
- Raftery, A. E. (1995). "Bayesian model selection in social research," in *Sociological Methodology 1995*, ed P. V. Marsden (Cambridge, MA: Blackwell), 111–196.
- Reed, H. C., Gemmink, M., Broens-Paffen, M., Kirschner, P. A., and Jolles, J. (2015). Improving multiplication fact fluency by choosing between competing answers. *Res. Math. Educ.* 17, 1–19. doi: 10.1080/14794802.2014.962074
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* 6, 461–464.
- Siegler, R. S. (1988). Individual differences in strategy choices: good students, not-so-good students, and perfectionists. *Child Dev.* 59, 833–851. doi: 10.2307/1130252
- Siegler, R. S. (1996). *Emerging Minds*. New York, NY: Oxford University Press.
- Siegler, R. S., and Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: testing predictions of ASCM using the choice/no-choice method. *J. Exp. Psychol.* 126:71. doi: 10.1037/0096-3445.126.1.71
- Siegler, R. S., and Stern, E. (1998). Conscious and unconscious strategy discoveries: a microgenetic analysis. *J. Exp. Psychol.* 127, 377–397. doi: 10.1037/0096-3445.127.4.377
- Smith-Chant, B. L., and LeFevre, J.-A. (2003). Doing as they are told and telling it like it is: self-reports in mental arithmetic. *Mem. Cognit.* 31, 516–528. doi: 10.3758/BF03196093
- Thevenot, C., Castel, C., Fanget, M., and Fayol, M. (2010). Mental subtraction in high- and lower skilled arithmetic problem solvers: verbal report versus operand-recognition paradigms. *J. Exp. Psychol.* 36, 1242–1255. doi: 10.1037/a0020447
- Torbeyns, J., Verschaffel, L., and Ghesquière, P. (2005). Simple addition strategies in a first-grade class with multiple strategy instruction. *Cogn. Instruc.* 23, 1–21. doi: 10.1207/s1532690xc2301_1
- Torbeyns, J., Verschaffel, L., and Ghesquière, P. (2006). The development of children's adaptive expertise in the number domain 20 to 100. *Cogn. Instruc.* 24, 439–465. doi: 10.1207/s1532690xc2404_2
- Torbeyns, J., De Smedt, B., Ghesquière, P., and Verschaffel, L. (2009a). Acquisition and use of shortcut strategies by traditionally schooled children. *Educ. Stud. Math.* 71, 1–17. doi: 10.1007/s10649-008-9155-z
- Torbeyns, J., Ghesquière, P., and Verschaffel, L. (2009b). Efficiency and flexibility of indirect addition in the domain of multi-digit subtraction. *Learn. Instruc.* 19, 1–12. doi: 10.1016/j.learninstruc.2007.12.002
- Torbeyns, J., and Verschaffel, L. (2016). Mental computation or standard algorithm? Children's strategy choices on multi-digit subtractions. *Eur. J. Psychol. Ed.* 31, 99–116. doi: 10.1007/s10212-015-0255-8
- Trbovich, P. L., and LeFevre, J. A. (2003). Phonological and visual working memory in mental addition. *Mem. Cognit.* 31, 738–745. doi: 10.3758/BF03196112
- Verschaffel, L., Torbeyns, J., De Smedt, B., Luwel, K., and Van Dooren, W. (2007). Strategy flexibility in children with mathematical difficulties. *Educ. Child Psychol.* 24, 16–27.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychon. Bull. Rev.* 14, 779–804. doi: 10.3758/BF03194105
- Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., and Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology: an empirical comparison using 855 *t* tests. *Perspect. Psychol. Sci.* 6, 291–298. doi: 10.1177/1745691611406923
- Widaman, K. F., Little, T. D., Geary, D. C., and Cormier, P. (1992). Individual differences in the development of skill in mental addition: internal and external validation of chronometric models. *Learn. Individ. Differ.* 4, 167–213.
- Xu, C., Wells, E., LeFevre, J. A., and Imbo, I. (2014). Strategic flexibility in computational estimation for Chinese- and Canadian-educated adults. *J. Exp. Psychol.* 40:1481. doi: 10.1037/a0037346
- Yee, T. W. (2015). *Vector Generalized Linear and Additive Models: With an Implementation in R*. New York, NY: Springer.
- Yee, T. W., and Wild, C. J. (1996). Vector generalized additive models. *J. R. Statist. Soc. Series B*, 58, 481–493.

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

Copyright © 2018 Caviola, Mammarella, Pastore and LeFevre. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Do Exact Calculation and Computation Estimation Reflect the Same Skills? Developmental and Individual Differences Perspectives

Dana Ganor-Stern*

Achva Academic College, Arugot, Israel

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

Patrick Lemaire,
Aix-Marseille Université, France
Fuchang Liu,
Wichita State University, United States

*Correspondence:

Dana Ganor-Stern
danaganor@gmail.com;
danaga@bgu.ac.il

Specialty section:

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

Received: 27 January 2018

Accepted: 09 July 2018

Published: 27 July 2018

Citation:

Ganor-Stern D (2018) Do Exact
Calculation and Computation
Estimation Reflect the Same Skills?
Developmental and Individual
Differences Perspectives.
Front. Psychol. 9:1316.
doi: 10.3389/fpsyg.2018.01316

Groups of children in 4th, 5th, and 6th grades and college students performed exact calculation and computation estimation tasks with two-digit multiplication problems. In the former they calculated the exact answer for each problem, and in the latter they estimated whether the result of each problem was larger or smaller than a given reference number. The analyses of speed and accuracy both showed different developmental patterns of the two tasks. While the accuracy of exact calculation increased with age in childhood, the accuracy of the estimation task reached its maximum level already in 4th grade and did not change with age. The reaction time of the exact calculation task was longer than that of the estimation task. The reaction time for both tasks remained constant in childhood and decreased in adulthood, with the improvement in speed larger for the exact calculation task. Similarly, within group variability in accuracy was larger in the exact calculation task than in the computation estimation task. Finally, low correlation was found between the accuracy of the two tasks. Together, these findings suggest that exact calculation and computation estimation reflect at least in part different skills.

Keywords: numerical cognition, computation estimation, exact calculation, development, multi-digit arithmetic, individual differences

INTRODUCTION

The present study focuses on the ability to solve multi-digit multiplication problems exactly and approximately. Children learn in school to solve arithmetic problems exactly. It has been shown that in the early stages of multiplication skill acquisition children use various calculation techniques to solve single digit (1D) multiplication problems (e.g., Siegler, 1988; Koshmider and Ashcraft, 1991). Similar to the process that occurs for single digit addition problems, with practice children gradually shift to solving such problems through retrieval from memory (e.g., Ashcraft and Battaglia, 1978; Ashcraft, 1992; LeFevre et al., 1996). Such a strategy shift is assumed to be due to an associative network stored in long term memory that includes the single digit multiplication or addition problems together with their respective answers (e.g., Siegler, 1988; Koshmider and Ashcraft, 1991). The formation of this associative network depends on an extensive practice with such problems and thus it is more likely to be formed when the problem set is small, as in the case of single digit multiplication or addition problems (e.g., Zbrodoff and Logan, 1986; Logan and Klapp, 1991).

Much less research was devoted to the investigation of how multidigit multiplication problems are solved (e.g., van der Ven et al., 2015). The number of multidigit numbers is substantially larger than the number of single digit numbers, and therefore the number of multiplication problems composed of such numbers is also greater. Each problem will thus receive less practice, and this will lead to weak, if any, associations between the problems and their respective answers (Siegler, 1988; Koshmider and Ashcraft, 1991). Thus, although the ability to solve complex multi-digit arithmetic problems via calculation improves with age during childhood and from childhood to adulthood, as indicated by increased accuracy and speed (Jordan et al., 2003; Ulf, 2010), such problems in large will not be solved in any stage via retrieval, but rather through the application of multi-step calculation algorithms, that rely heavily on working memory (Lee and Kang, 2002; DeStefano and LeFevre, 2004; Kalamian and Lefevre, 2007; Raghubar et al., 2010). Due to working memory limitations people often use paper and pencil, or even turn to calculators.

In many real life circumstances it is sufficient to produce approximate, rather than exact answers to complex arithmetic problems such as multi-digit multiplication problems. For example, when planning a wedding party one might be interested in the approximate rather than exact cost involved in inviting 130 people, with the price of catering being 27\$ per person. The process of producing an approximate answer to an arithmetic problem is called computation estimation (Rubinstein, 1985). Its main advantage is that it takes less time and attentional resources than exact calculation, and thus can be used in circumstances where time or attention resources are limited. It should be noted that the importance of computation estimation is not undermined by the wide use of calculators, as using a calculator to solve a multidigit problem is prone to typing errors, and computation estimation can be used as a sanity check to quickly evaluate whether the calculator generated answer is reasonable (LeFevre et al., 1993; Siegler and Booth, 2005).

Despite its importance, computation estimation has received relatively little attention in the educational system and in the numerical cognition literature (e.g., Siegler and Booth, 2005). One way to investigate this skill is using the estimation production task, in which participants are asked to produce an approximated answer for an arithmetic problem (e.g., LeFevre et al., 1993; Dowker et al., 1996; Lemaire and Lecacheur, 2002; Imbo and LeFevre, 2011). It has been shown that the accuracy of such approximated answer improves with age, although it is still poor even for adults (e.g., LeFevre et al., 1993). An examination of the strategies used based on participants self reports, reveals the use of various rounding techniques (e.g., Lemaire and Lecacheur, 2002; Lemaire et al., 2004; Siegler and Booth, 2005). With age there is an increased use of the more complex rounding strategies, and more adaptivity in strategy selection, such that the rounding procedure that introduced the least amount of error is chosen more often (e.g., Lemaire and Lecacheur, 2011). With age there is also more frequent use of post-compensation procedures to correct for the error introduced by the rounding procedures (e.g., Siegler and Booth, 2005).

In the estimation comparison task, another experimental paradigm used to study computation estimation in the context of multi-digit arithmetic, a multidigit multiplication problem was presented together with a reference number, and participants were required to estimate whether the exact answer to the problem was larger or smaller than the reference number (Ganor-Stern, 2015, 2016, 2017; Ganor-Stern and Weiss, 2015). The reference number was either far or close to the exact answer. The advantage of this task is that it enables the use of two distinct strategies. The first is the *approximated calculation strategy*, which involves rounding procedures, is the strategy mainly used in the estimation production task. The second is the *sense of magnitude strategy*, which involves an intuitive sense of magnitude without any calculation and it can be used only in this task due to the presence of the reference number. This strategy probably reflects the life-long experience with solving arithmetic problems and even the practice provided by the experimental session (Ganor-Stern and Weiss, 2015; Ganor-Stern, 2016). Past research has repeatedly shown adaptivity in strategy choice, as the approximated calculation strategy was used more often when the reference number was close to the exact answer, and thus the sense of magnitude cannot guarantee a correct response, while the sense of magnitude was used more often when the reference number was far from it. In terms of speed and accuracy it has been consistently shown that performance in this task is enhanced for reference numbers that are far vs. close to the exact answer, and for those that are smaller vs. larger than the exact answer (Ganor-Stern, 2015, 2016, 2017; Ganor-Stern and Weiss, 2015).

Ganor-Stern (2016) has investigated the developmental pattern of performance in this task looking at 4th graders, 6th graders and college students. There was some improvement in accuracy with age, as percent error was 22% for 4th graders and 17% for 6th graders and 16% for adults. This improvement was limited to trials in which the reference numbers were close to the exact answer; there was no improvement in accuracy for trials where the reference numbers were far from the exact answer. There was a substantial improvement in speed with age, especially in adulthood. Thus, while 4th and 6th graders took on average 12 and 11 s to respond, respectively, adults responded in only 4 s. In terms of strategy use, with age there was a decrease in the use of the sense of magnitude strategy and an increase in the use of the approximated calculation strategy, which presumably underlies the improvement in the accuracy for the close reference trials.

Present Study

Despite the fact that estimation of the results of arithmetic problems is a useful skill in life it is still debated whether it reflects the same skill as solving the same problems exactly. This is the main question addressed by the current research. Research conducted on young children (between ages 5 and 9 years old) has shown positive relationship between the exact calculation and the estimation skills using addition problems (e.g., Dowker, 1997). Although when looking at children who show especially weak exact calculation skills, their estimates were found to be similar to those with average calculation skills, which implies

some dissociation between the two skills (Dowker, 2005). In a similar manner, a study by Liu (2013) conducted on adults has shown that the problem size affected exact calculation but not estimation. Specifically, the larger the problem the higher the error rate and reaction time when participants solved it exactly, but not when they estimated its answer.

The present study expands past research by using a different estimation task than Dowker (1997) and Liu (2013), by looking at the developmental patterns of the estimation and exact calculation tasks from childhood to adulthood, at individual differences within each task and age group, and at the correlation between performance in the estimation and exact calculation tasks.

Specifically, groups of 4th graders, 5th graders, 6th graders, and college students solved a set of 20 2D multiplication problems exactly, and estimated the results of another set of 40 similar 2D multiplication problems relative to a reference number using the estimation comparison task (Ganor-Stern, 2015, 2016, 2017). In both tasks, speed and accuracy were analyzed by age. For the estimation task the analysis looked also at the effects of the reference number characteristics (its magnitude relative to the exact answer and its distance from the exact answer) on performance.

As to the predictions, on the one hand, one might expect a strong relationship between performance in the two tasks, as they both require arithmetic processing of the same stimuli (e.g., Dowker, 1997). On the other hand, past research provided evidence for dissociations between exact calculation and approximation, as exact calculation is language-dependent while approximation is not (e.g., Pica et al., 2004). Moreover, they seem to activate different areas in the brain. During exact calculation there is strong activation in the left inferior prefrontal cortex, while during approximation there is activation in the inferior parietal lobule in both hemispheres (e.g., Dehaene et al., 1999).

Furthermore, while the exact calculation task used in the current study involves a long working-memory-dependent algorithmic process, the computation estimation task seems to reflect a basic sense of magnitude together with a shortened algorithmic process (Ganor-Stern and Weiss, 2015; Ganor-Stern, 2016). Indeed, the results of a recent study on the effect of attention deficit hyperactivity disorder (ADHD) on estimation vs. exact calculation support some dissociation between exact calculation and the two strategies involved in the estimation comparison task. Participants with ADHD, which is assumed to involve working memory and executive function deficiencies (Castellanos et al., 2006), were impaired when conducting exact calculation and when using the approximated calculation strategy in the estimation task, but not when the sense of magnitude strategy was used (Ganor-Stern and Steinhorn, 2018).

As to development with age, based on past research that showed little improvement in estimation accuracy from childhood to adulthood (Ganor-Stern, 2016), but a significant improvement in the accuracy of exact calculation (e.g., Ulf, 2010) we expect to see more improvement with age in the accuracy of exact compared to approximated calculation. Speed is expected to increase in both tasks, although to a greater extent in the

exact calculation task (e.g., Ulf, 2010). Finally, we expect to find more variability in performance (in accuracy or speed) across participants within each age group in the exact calculation compared to the estimation task (Dowker, 2005).

MATERIALS AND METHODS

Participants

There were four groups of participants. Thirty three children from fourth grade (20 females), 33 children from 5th grade (16 females), 33 children from 6th grade (18 females), and 25 college students (23 females). The children were from three public schools in the center of Israel, and the college students were from a public academic college. The average age of the 4th graders was 9.8 years old, of the 5th graders it was 10.9 years old, of the 6th graders it was 12.04 years old, and of the college students it was 23.1 years old.

Ethics Statement

The procedure was approved by the ethics committees of the Israeli Ministry of Education and of Achva Academic College, Israel. The college students provided written informed consent to participate in this study. Adhering to the policy of the Ministry of Education IRB, the parents of the school children denied consent by returning an enclosed form.

Stimuli

The stimuli were 60 2-digit (D) multiplication problems. The problems in the estimation and in the exact calculation tasks were different, however, they were constructed with the same following restrictions. There were no tie problems. No operand had 0 as units digit. No reversed orders of operands were used (43×76 was not used with 76×43). The larger operand was on the left in half of the problems, and on the right in the other half. The problems for the estimation task were taken from Ganor-Stern (2016). The range of exact answers in the exact calculation task was 903–6391, and in the estimation task it was 768–8178.

The multiplication problems in both tasks were printed on sheets of paper. The exact calculation task included two sets of 10 problems each. Four problems were printed vertically on each page to leave space for the calculation. The estimation task that included 40 items was printed on a booklet. Each item was composed of a 2D multiplication problem with a reference number below it, and the word “smaller” written beneath the reference number on the left side, and the word “larger” written on the right side (Ganor-Stern, 2016). Four problems were printed on a sheet of paper. The reference numbers were of 4 types: (1) one which was about one fifth of the exact answer, (2) one which was about five times the exact answer, (3) one which was about one half of the exact answer, and (4) one which was about twice the exact answer. Ten problems were associated with each reference number type. Types (1) and (2) are the far condition, and types (3) and (4) are the close condition. In (1) and (3) the exact answer is larger than the reference number, and in (2) and (4) the exact answer is smaller than the reference number.

Reference numbers were rounded to the nearest hundred. In half of the trials the exact answer was larger than the reference number, and in the other half it was smaller than the reference number.

Procedure

The experiment took place in a class setting. The experimenter explained the participants that they will be solving 2D multiplication problems. The participants were first given a set of 10 2D multiplication problems printed on sheets of paper, and were instructed to solve the problems exactly on the paper sheets. Then they were given a booklet with 40 estimation items. Each item consisted of a 2D multiplication problem with a reference number below it, and the word “smaller” on the left side, and the word “larger” on the right side. The participants were asked to indicate for each problem whether they estimated the exact answer to be smaller or larger than the given reference number by marking either the word “larger” or the word “smaller.” Finally, the participants were given a new set of 10 2D multiplication problems printed on sheets of paper to solve them exactly on the paper. There were no time limits. For each task, the experimenters documented on each participant’s sheet of paper the time he/she started each set of problems. The students were asked to raise their hands when they finished the current set. The experimenter documented the time the participant ended the task on the paper sheet, and handed him/her the following set. Participants were not allowed to use calculators in any of the tasks.

RESULTS

The performance measures for each task were the accuracy for each problem and the solution time for each problem set, which was divided by the number of problems, for an average solution time for a single problem. The analyses examined the developmental patterns within each task, the between-participants variability within each task, and the relationship between performance in the two tasks.

The Developmental Pattern in the Exact Calculation Task

A one way Analysis of Variance (ANOVA) on the percentage of correct responses in the exact calculation task with age as a between-participants variable has shown that percent of correct responses increased with age ($F_{3,118} = 8.07$, $MSE = 8.49$, $p = 0.0001$, $\eta_p^2 = 0.17$). Sheffe *post hoc* tests showed that 4th graders were less accurate (36%) than the other groups ($p < 0.05$), that did not differ (Figure 1). Percent of correct responses was 62, 69, and 62 for 5th graders, 6th graders and adults, respectively. The speed analysis revealed a significant effect of age ($F_{3,120} = 24.40$, $MSE = 1386.7$, $p = 0.0001$, $\eta_p^2 = 0.38$). Sheffe *post hoc* tests showed that the adults were faster than the children groups ($p < 0.05$), that did not differ. Average response time was 88.82, 90.00, and 89.46 s for the 4th graders, 5th graders, and 6th graders, respectively, and it was 18.12 s for adults (Figure 2).

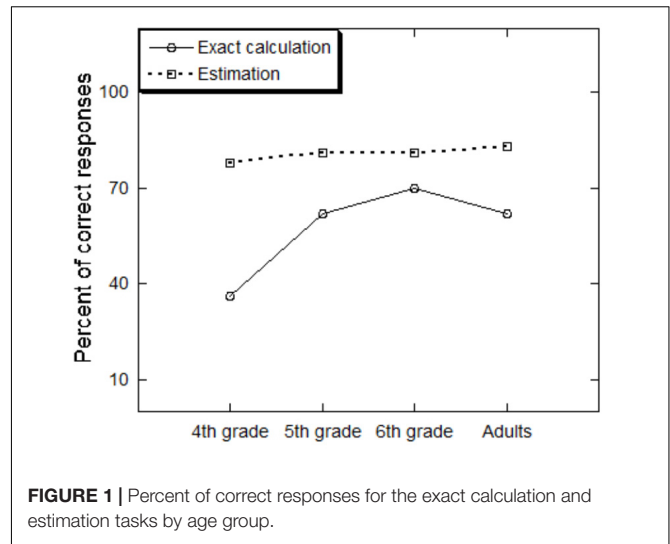


FIGURE 1 | Percent of correct responses for the exact calculation and estimation tasks by age group.

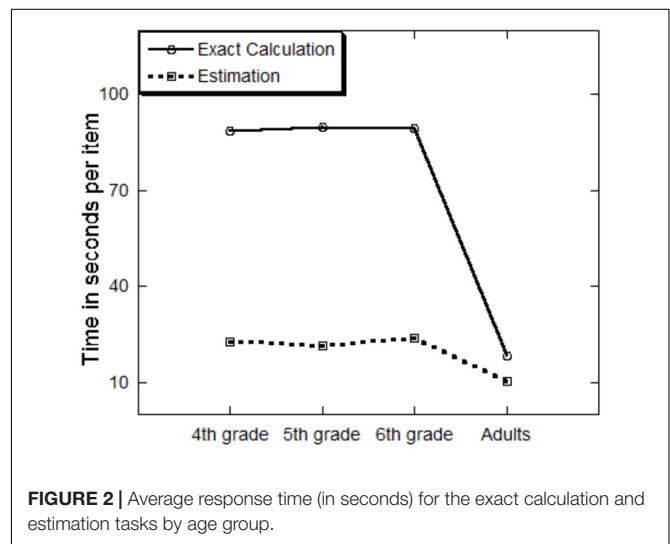


FIGURE 2 | Average response time (in seconds) for the exact calculation and estimation tasks by age group.

The Developmental Pattern in the Estimation Task

An ANOVA on the average response time has shown a significant effect of age ($F_{3,119} = 9.72$, $MSE = 48.52$, $p = 0.0001$, $\eta_p^2 = 0.20$). Again sheffe *post hoc* tests have shown that adults were faster than the children groups when solving the estimation task ($p < 0.01$), while the children groups did not differ in speed (Figure 2). Thus, while it took adults on average 10.2 s to respond to a problem, it took 4th graders about 22.73 s, 5th graders 21.5 s, and 6th graders 23.77 s.

The accuracy analysis included in addition to the age factor also the within-participant factors of the size of the reference number (larger vs. smaller than the exact answer) and its distance from the exact answer (far vs. close).¹ As was found in past

¹ As the estimation task included only one problem set, we had an accuracy measure for each item and a speed measure for the whole set. This enabled us to analyze the effect of reference number characteristics and to calculate the split half reliability of the estimation task for accuracy only, and not for speed.

research with the same task (Ganor-Stern, 2015, 2016, 2017; Ganor-Stern and Weiss, 2015), accuracy was higher when the reference number was far (83%) compared to close (78%) to the exact answer ($F_{1,120} = 19.66$, $MSE = 0.36$, $p = 0.0001$, $\eta_p^2 = 0.14$). It was also higher when the reference number was smaller (83%) than the exact answer compared to when it was larger (79%) than it, although the effect was marginally significant ($F_{1,120} = 3.55$, $MSE = 0.06$, $p = 0.06$, $\eta_p^2 = 0.03$). Importantly, as can be seen in **Figure 1**, accuracy did not differ across the age groups ($F < 1$).

Cross-Participants Variability in Performance in the Exact Calculation and Estimation Tasks

To examine cross-participants variability in performance the coefficient of variability was calculated for each age group and for each task, for accuracy and speed separately. This was done by dividing the standard deviation of accuracy and of speed across participants by the group average and multiplying by 100. The results (**Table 1**) show that the coefficient of variability in accuracy was higher for the exact calculation task compared to the computation estimation task, and that it decreased with age for the former but not for the latter. The coefficient of variability in speed does not show a consistent pattern across tasks or across age.

Relationship Between the Performance in the Exact Calculation and Estimation Tasks

To examine the relationship between the two tasks, we calculated the correlation between the accuracy of the two tasks and the reaction time of the two tasks. This was done separately for each age group, and across age groups (**Table 2**). The correlation between the accuracy of the two tasks, collapsed across the age groups, was 0.35 ($p < 0.05$), and between the speed of the two tasks was 0.60 ($p < 0.05$). As can be seen in **Table 2**, the pattern of stronger inter-task correlation in speed than in accuracy was found in most age groups. This is possibly due to the low variability in accuracy found in the estimation comparison task. Accuracy level in the estimation task showed the least variability across the age groups (**Figure 1**) and across-participants within each age group (**Table 1**).

To examine whether the low inter-task correlation (at least in accuracy) is due to low reliability of the tasks, we calculated split

TABLE 2 | Inter-task correlation and reliability coefficients by task and age group.

	Inter-task correlation		Split-half reliability		
	Accuracy	Speed	Exact Calculation Accuracy	Exact Calculation Speed	Estimation Accuracy
All	0.35	0.60	0.87	0.88	0.79
4th grade	0.35	0.38	0.77	0.86	0.58
5th grade	0.36	0.62	0.88	0.78	0.89
6th grade	0.13	0.35	0.83	0.95	0.79
Adults	0.53	-0.07	0.89	0.92	0.84

The numbers in plain font represent significant correlations ($p < 0.05$), while the numbers in italics and in a smaller font represent insignificant correlations.

half reliabilities for each of the tasks. As can be seen in **Table 2** the split half reliabilities of the two tasks were relatively high (in most cases they were higher than 0.80), thus suggesting that the inter-task correlations were not restricted by the tasks reliabilities¹.

DISCUSSION

From a developmental perspective, accuracy in the exact calculation task improved from 4th to 5th grade and then remained unchanged up to adulthood. Note that percent of correct responses hardly reached 70%, far from perfect accuracy, thus suggesting that participants even in adulthood are not proficient in solving multi-digit multiplication problems, probably due to the wide use of calculators. Note that the accuracy level of the two tasks is not comparable as the exact calculation task is an open ended task, while the computation estimation task is a forced choice one. Thus, what seems to be informative is the different patterns across age. While exact calculation accuracy increased with age, accuracy of the computation estimation task did not change by age at all. As to speed, speed improved in both tasks mainly in adulthood, although the increase was much more pronounced for the exact calculation task.

Past research has found a continuous improvement in accuracy (van der Ven et al., 2015) and in speed (Koshmider and Ashcraft, 1991; De Brauwer and Fias, 2009) throughout childhood when solving single digit multiplication problems exactly. In the present study the only improvement in accuracy of exact calculation was found between 4th and 5th grades. The reason might lie in the difference between single vs. multiple digit multiplication. Single digit multiplication is practiced on its own, and as part of multidigit multiplication, and thus it continues to improve. Multidigit multiplication is practiced much less, in part due to the increased use of calculators.

The main improvement in speed is seen in adulthood. This is true for both tasks but it is more apparent for the exact calculation task. This improvement could be due to improvement of domain-specific skills, such as calculation skills (e.g., Pauli et al., 1994), or domain general factors, such working memory and decision processes (e.g., Berg, 2008). The fact that the improvement in speed in adulthood was not accompanied with an improvement

TABLE 1 | Coefficient of variability in accuracy and speed by task and age group.

Age group	Accuracy		Speed	
	Exact calculation	Computation estimation	Exact calculation	Computation estimation
4th grade	77.72	15.64	47.73	44.99
5th grade	49.61	17.71	45.13	66.87
6th grade	35.01	20.00	44.14	40.69
Adults	51.44	17.79	38.32	36.02

in accuracy might suggest that domain general factors accounted for it.

Note that the accuracy levels in the estimation task of the current study are comparable to those found in past research using the same task and the same age groups (Ganor-Stern, 2016). The reaction time here are longer due to the use of paper and pencil, however, the speed patterns are similar. In both studies speed remained unchanged in childhood and it improved considerably in adulthood. The facts that similar patterns were found in the two studies both in accuracy and in speed despite the use of different procedures [i.e., the current study used a paper and pencil procedure, while in Ganor-Stern (2016) the experiment was computerized] provide convergent validity to the current results.

The different developmental trajectories of the two tasks suggest that they do not reflect the same skill. In a consistent manner, the analysis of variability has shown that the variability in accuracy was smaller for the computation estimation task than for the exact calculation task. Moreover, while for the exact calculation task this variability decreased with age, consistent with past research (De Brauwer and Fias, 2009), for the computation estimation task it did not. The relatively low correlation between the accuracy of the two tasks also corroborates the dissociation between the two tasks.

The present research showing an increase in accuracy between 4th and 6th grades in the exact calculation task is in line with Ulf (2010), who found a similar pattern. Note, however, that in contrast to the present findings, Ulf also reported an improvement in approximated calculation between 4th and 6th grades. A possible explanation for this difference is the nature of the estimation task used. In Ulf (2010) children were given addition and subtraction problems composed of 2D numbers presented vertically. Each problem was accompanied with two proposed answers, and the task was to choose the answer that was closest to the correct answer. Such a task might encourage participants to solve the problem exactly, and thus might show similar improvements with age for the exact calculation and estimation tasks.

The current estimation comparison task seems to capture not only approximated calculation but also sense of magnitude for the results possible for such multidigit multiplication problems. This is indicated by the use of sense of magnitude strategy itself, and by the adaptive choice between the sense of magnitude and the approximated calculation strategies. This sense of magnitude might be related to the Approximate Number System (ANS), which represents magnitudes in an approximated manner, develops early, and is language independent (Ansari, 2008; Mazzocco et al., 2011a,b; Park and Brannon, 2013).

Across studies and age groups participants use the approximated calculation strategy more often when the reference numbers are close to the exact answer than when they are from it, suggesting that participants have a rough sense for how big the answers could be, and thus use the approximated calculation strategy more often when the reference number is within this range and the sense of magnitude strategy when the

reference number it is outside of it (Ganor-Stern, 2015, 2016, 2017; Ganor-Stern and Weiss, 2015). Importantly, this pattern of adaptive strategy choice was found even for children as young as 4th graders (Ganor-Stern, 2016) and for adults diagnosed with developmental dyscalculia (Ganor-Stern, 2017).

The conclusion of the current study that exact calculation and estimation do not reflect the same skill is consistent with past research that argue for a dissociation between estimation and exact calculation (e.g., Pica et al., 2004; Liu, 2013), more generally it is compatible with theories that emphasize the componential nature of arithmetic (e.g., Dowker, 2005).

The current study did not collect information about strategy use. Future research, in which participants describe the strategy they used on a trial by trial basis should look at the relationship between exact calculation and estimation performance separately for the two strategies used. It is predicted that the correlation between the accuracy of the estimation and exact calculation tasks will be higher for trials in which the approximated calculation strategy was used.

Limitations of the Present Study

The fact that the computation estimation task was a forced choice task, and the exact calculation task was an open ended one prevents a direct comparison of the accuracy and speed of the two tasks, and this might be seen as a limitation of the current study. The rationale for this design is that the use of a forced choice task with reference numbers in the estimation task allowed using sense of magnitude when solving this task, especially with far reference numbers. For the exact calculation task, the use of a forced choice task might have encouraged participants to use shortcut strategies, such as parity rules, (e.g., Lemaire and Fayol, 1995) rather than to go through the whole solution process, and thus an open ended format was used. Furthermore, the task order was determined by school considerations, which did not allow for a random or counterbalanced design. As a consequence no conclusions on the effect of performing one task on the other task can be drawn. Finally, the measurement of speed was possible for the whole set rather than for each item, due to the use of paper and pencil, rather than a computerized task. This was done because solving complex multidigit multiplication problems is usually done in everyday life with paper and pencil, and the experimental setting tried to mimic these natural conditions.

AUTHOR CONTRIBUTIONS

DG-S conception and design, statistical analyses, interpretation of data, and drafting the manuscript.

ACKNOWLEDGMENTS

The author thanks Korin Avital, Vered Malki, Lital Danino, and Avital Moshinski for their help in data collection and analysis, and for the participants who took part in this study.

REFERENCES

- Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nat. Rev. Neurosci.* 9, 278–291. doi: 10.1038/nrn2334
- Ashcraft, M. H. (1992). Cognitive arithmetic: a review of data and theory. *Cognition* 44, 75–106. doi: 10.1016/0010-0277(92)90051-I
- Ashcraft, M. H., and Battaglia, J. (1978). Cognitive arithmetic: evidence for retrieval and decision processes in mental addition. *J. Exp. Psychol. Hum. Learn. Mem.* 4, 527–538. doi: 10.1037/0278-7393.4.5.527
- Berg, D. H. (2008). Working memory and arithmetic calculation in children: the contributory roles of processing speed, short-term memory, and reading. *J. Exp. Child Psychol.* 99, 288–308. doi: 10.1016/j.jecp.2007.12.002
- Castellanos, F. X., Sonuga-Barke, E. J., Milham, M. P., and Tannock, R. (2006). Characterizing cognition in ADHD: beyond executive dysfunction. *Trends Cogn. Sci.* 10, 117–123. doi: 10.1016/j.tics.2006.01.011
- De Brauwier, J., and Fias, W. (2009). A longitudinal study of children's performance on simple multiplication and division problems. *Dev. Psychol.* 45, 1480–1496. doi: 10.1037/a0015465
- Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., and Tsivkin, S. (1999). Sources of mathematical thinking: behavioral and brain-imaging evidence. *Science* 284, 970–974. doi: 10.1126/science.284.5416.970
- DeStefano, D., and LeFevre, J. (2004). The role of working memory in mental arithmetic. *Eur. J. Cogn. Psychol.* 16, 353–386. doi: 10.1080/09541440244000328
- Dowker, A. (1997). Young children's addition estimates. *Math. Cogn.* 3, 141–154. doi: 10.1080/135467997387452
- Dowker, A. (2005). *Individual Differences in Arithmetic: Implications for Psychology, Neuroscience and Education*. Hove: Psychology Press. doi: 10.4324/9780203324899
- Dowker, A., Flood, A., Griffiths, H., Harriss, L., and Hook, L. (1996). Estimation strategies of four groups. *Math. Cogn.* 2, 113–135. doi: 10.1080/135467996387499
- Ganor-Stern, D. (2015). When you don't have to be exact – Investigating computation estimation skills with a comparison task. *Acta Psychol.* 154, 1–9. doi: 10.1016/j.actpsy.2014.10.010
- Ganor-Stern, D. (2016). Solving math problems approximately: a developmental perspective. *PLoS One* 11:e0155515. doi: 10.1371/journal.pone.0155515
- Ganor-Stern, D. (2017). Can dyscalculics estimate the results of arithmetic problems? *J. Learn. Disabil.* 50, 23–33. doi: 10.1177/0022219415587785
- Ganor-Stern, D., and Steinhorn, O. (2018). ADHD and math - The differential effect on calculation and estimation. *Acta Psychol.* 188, 55–64. doi: 10.1016/j.actpsy.2018.05.013
- Ganor-Stern, D., and Weiss, N. (2015). Tracking practice effects in computation estimation. *Psychol. Res.* 80, 434–448. doi: 10.1007/s00426-015-0720-7
- Imbo, I., and LeFevre, J.-A. (2011). Cultural differences in strategic behavior: a study in computational estimation. *J. Exp. Psychol. Learn. Mem. Cogn.* 37, 1294–1301. doi: 10.1037/a0024070
- Jordan, N. C., Hanich, L. B., and Kaplan, D. (2003). Arithmetic fact mastery in young children: a longitudinal investigation. *J. Exp. Child Psychol.* 85, 103–119. doi: 10.1016/S0022-0965(03)00032-8
- Kalaman, D., and Lefevre, J.-A. (2007). Working memory demands of exact and approximate addition. *Eur. J. Cogn. Psychol.* 19, 187–212. doi: 10.1080/09541440600713445
- Koshmider, J. W., and Ashcraft, M. H. (1991). The development of children's mental multiplication skills. *J. Exp. Child Psychol.* 51, 53–89. doi: 10.1016/0022-0965(91)90077-6
- Lee, K.-M., and Kang, S.-Y. (2002). Arithmetic operation and working memory: differential suppression in dual tasks. *Cognition* 83, B63–B68. doi: 10.1016/S0010-0277(02)00010-0
- LeFevre, J.-A., Bisanz, J., Daley, K. E., Buffone, L., Greenham, S. L., and Sadesky, G. S. (1996). Multiple routes to solution of single-digit multiplication problems. *J. Exp. Psychol. Gen.* 125, 284–306. doi: 10.1037/0096-3445.125.3.284
- LeFevre, J.-A., Greenham, S. L., and Waheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. *Cogn. Instr.* 11, 95–132. doi: 10.1207/s1532690xci1102-1
- Lemaire, P., Arnaud, L., and Lecacheur, M. (2004). Adults' age-related differences in adaptivity of strategy choices: evidence from computational estimation. *Psychol. Aging* 19, 467–481. doi: 10.1037/0882-7974.19.3.467
- Lemaire, P., and Fayol, M. (1995). When plausibility judgments supersede fact retrieval: the example of the odd-even effect on product verification. *Mem. Cogn.* 23, 34–48. doi: 10.3758/BF03210555
- Lemaire, P., and Lecacheur, M. (2002). Children's strategies in computational estimation. *J. Exp. Child Psychol.* 82, 281–304. doi: 10.1016/S0022-0965(02)00107-8
- Lemaire, P., and Lecacheur, M. (2011). Age-related changes in children's executive functions and strategy selection: a study in computational estimation. *Cogn. Dev.* 26, 282–294. doi: 10.1016/j.cogdev.2011.01.002
- Liu, F. (2013). Are exact calculation and computational estimation categorically different? *Appl. Cogn. Psychol.* 27, 672–682. doi: 10.1002/acp.2947
- Logan, G. D., and Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *J. Exp. Psychol. Learn. Mem. Cogn.* 17, 179–195. doi: 10.1037/0278-7393.17.2.179
- Mazzocco, M. M., Feigenson, L., and Halberda, J. (2011a). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Dev.* 82, 1224–1237. doi: 10.1111/j.1467-8624.2011.01608.x
- Mazzocco, M. M., Feigenson, L., and Halberda, J. (2011b). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS One* 6:e23749. doi: 10.1371/journal.pone.0023749
- Park, J., and Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychol. Sci.* 24, 2013–2019. doi: 10.1177/0956797613482944
- Pauli, P., Lutzenberger, W., Rau, H., Birbaumer, N., Rickard, T. C., Yaroush, R. A., et al. (1994). Brain potentials during mental arithmetic: effects of extensive practice and problem difficulty. *Cogn. Brain Res.* 2, 21–29. doi: 10.1016/0926-6410(94)90017-5
- Pica, P., Lemer, C., Izard, V., and Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306, 499–503. doi: 10.1126/science.1102085
- Raghubar, K. P., Barnes, M. A., and Hecht, S. A. (2010). Working memory and mathematics: a review of developmental, individual difference, and cognitive approaches. *Learn. Individ. Differ.* 20, 110–122. doi: 10.1016/j.lindif.2009.10.005
- Rubinstein, R. (1985). Computational estimation and related mathematical skills. *J. Res. Math. Educ.* 16, 106–119. doi: 10.2307/748368
- Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *J. Exp. Psychol. Gen.* 117, 258–275. doi: 10.1037/0096-3445.117.3.258
- Siegler, R. S., and Booth, J. (2005). "Development of numerical estimation: A review," in *Handbook of Mathematical Cognition*, ed. J. Campbell (New York, NY: Psychology Press), 197–212.
- Ulf, T. (2010). Skill development in different components of arithmetic and basic cognitive functions: findings from a 3-year longitudinal study of children with different types of learning difficulties. *J. Educ. Psychol.* 102, 115–134. doi: 10.1037/a0016838
- van der Ven, S. H., Straatemeier, M., Jansen, B. R., Klinkenberg, S., and van der Maas, H. L. (2015). Learning multiplication: an integrated analysis of the multiplication ability of primary school children and the difficulty of single digit and multidigit multiplication problems. *Learn. Individ. Differ.* 43, 48–62. doi: 10.1016/j.lindif.2015.08.013
- Zbrodoff, N. J., and Logan, G. D. (1986). On the autonomy of mental processes: a case study of arithmetic. *J. Exp. Psychol. Gen.* 115, 118–130. doi: 10.1037/0096-3445.115.2.118

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

Copyright © 2018 Ganor-Stern. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Variability in Single Digit Addition Problem-Solving Speed Over Time Identifies Typical, Delay and Deficit Math Pathways

Robert A. Reeve^{1*}, Sarah A. Gray¹, Brian L. Butterworth^{1,2} and Jacob M. Paul¹

¹ Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, VIC, Australia, ² Centre for Educational Neuroscience, University College London, London, United Kingdom

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

Annemie Desoete,
Ghent University, Belgium
Julie Ann Jordan,
Queen's University Belfast,
United Kingdom
Jennifer M. Zosh,
Pennsylvania State University,
United States

*Correspondence:

Robert A. Reeve
r.reeve@unimelb.edu.au

Specialty section:

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

Received: 20 February 2018

Accepted: 30 July 2018

Published: 14 August 2018

Citation:

Reeve RA, Gray SA, Butterworth BL
and Paul JM (2018) Variability
in Single Digit Addition
Problem-Solving Speed Over Time
Identifies Typical, Delay and Deficit
Math Pathways.
Front. Psychol. 9:1498.
doi: 10.3389/fpsyg.2018.01498

We assessed the degree to which the variability in the time children took to solve single digit addition (SDA) problems longitudinally, predicted their ability to solve more complex mental addition problems. Beginning at 5 years, 164 children completed a 12-item SDA test on four occasions over 6 years. We also assessed their (1) digit span, visuospatial working memory, and non-verbal IQ, and (2) the speed with which they named single numbers and letters, as well the speed enumerating one to three dots as a measure of subitizing ability. Children completed a double-digit mental addition test at the end of the study. We conducted a latent profile analysis to determine if there were different SDA problem solving response time (PRT) variability patterns across the four test occasions, which yielded three distinct PRT variability patterns. In one pattern, labeled a typical acquisition pathway, mean PRTs were relatively low and PRT variability diminished over time. In a second pattern, labeled a delayed pathway, mean PRT and variability was high initially but diminished over time. In a third pattern, labeled a deficit pathway, mean PRT and variability remained relatively high throughout the study. We investigated the degree to which the three SDA PRT variability pathways were associated with (1) different cognitive ability measures, and (2) double-digit mental addition abilities. The deficit pathway differed from the typical and delayed pathway on the subitizing measure only, but not other measures; and the latter two pathways also differed from each other on the subitizing but not other measures. Double-digit mental addition problem solving success differed between each of the three pathways, and mean PRT variability differed between the typical and the delayed and deficit pathways. The latter two pathways did not differ from each other. The findings emphasize the value of examining individual differences in problem-solving PRT variability longitudinally as an index of math ability, and highlight the importance of subitizing ability as a diagnostic index of math ability/difficulties.

Keywords: typical, delayed, deficit math pathways, single digit addition problem solving speed variability, subitizing ability, longitudinal analysis

INTRODUCTION

One goal of early math instruction is to help children acquire the basic arithmetic skills necessary to solve more complex calculation problems. Ensuring children acquire good single digit addition (SDA) number fact abilities, for example, is a learning objective in many countries (OECD, 2014). While instructional emphases differ (e.g., from a focus on rote learning to reasoning strategies),

children tend to use so-called procedural strategies (e.g., counting all items) before so-called conceptual strategies (e.g., decomposition of number facts) to solve SDA problems (Butterworth, 2005; Geary and Hoard, 2005; Siegler, 2016); and, children may use both procedural and conceptual strategies on a single test occasion. While the association between the strategies used to solve SDA problems and problem-solving success varies within and across age, most children solve SDA problems eventually (Paul and Reeve, 2016). Nevertheless, this acquisition variability raises the possibility that different SDA acquisition pathways are embedded within a general acquisition pathway. Insofar as different SDA acquisition pathways can be identified, it is possible they lead to a single ability end-point (equifinality); it is also possible that different pathways reflect different ability profiles, which would have implications for our understanding of math development.

The present study addressed the issue of whether it is possible to distinguish typical, deficit and delay SDA acquisition pathways in primary-aged school children based on changes in the variability of the speed with which children solved SDA problems over a 5-year period. We focused on variability in SDA problem-solving speed because arguably it represents an index of changes in SDA problem-solving efficiency, especially when examined over time. Focusing on problem solving speed also allowed us to examine SDA problem solving after children were able to solve problems correctly. In general, it would be expected that children's SDA problem-solving speed trajectory would decline and become less variable over time. It is possible that problem-solving speed will decline slowly for some children (a delayed pathway?), or continues to be variable (a deficit pathway?) over time. Given the importance of SDA abilities in curricula, understanding the factors associated with different SDA developmental pathways may have diagnostic significance, as well as contribute insights to our understanding of the nature individual differences in math development more generally.

SDA Strategy Change

The strategies children employ to solve single addition problems skills, on average, change in their conceptual sophistication over time and are claimed to represent changes in math reasoning abilities (Baroody, 2003; Butterworth, 2005; Siegler, 2006; Geary et al., 2007; Jordan et al., 2009; Paul and Reeve, 2016). Children initially guess answers, following which they may use a *count all* strategy to individually enumerate the numbers of the two addends. Subsequently, they may adopt a *count on* strategy (specifying the cardinal value of the first addend, and sequentially enumerating the numbers of the second addend). Children may then employ a *min* strategy (counting on from the larger of the two addends when it is the second term). In time, they begin using more sophisticated strategies, including the *decomposition* of number facts and *retrieval* of answers from memory (Baroody and Dowker, 2003; Geary et al., 2007).

How should these changes in the acquisition of SDA problem solving abilities be characterized? As Siegler notes, the development of children's reasoning strategies is more variable than often acknowledged (Siegler, 2007, 2016). Siegler (1996, 2000) characterized reasoning development in terms

of changes in the selection of strategy options over time. Commonly, children use a mix of strategies to solve problems, with a progressive reduction of less sophisticated strategies accompanying the acquisition of problem solving ability (usually across age). That is, with age and/or experience, children solve problems more quickly and select more efficient strategies, and less efficient strategies disappear from their repertoire (Torbeyns et al., 2004).

Is strategy change the same for all children, or are there different strategy change profiles and, if there are, what do they imply about children's abilities? Siegler's overlapping wave model suggests the acquisition of problem solving competence may be analyzed along five dimension of change—path, rate, breadth, sources and variability (see Siegler, 2005). Siegler and colleagues (Siegler, 2005, 2016) suggest these dimensions may be studied using the so-called microgenetic method in which multiple observations of strategy change are made from the beginning of change to the point at which strategy-use becomes relatively stable. Strategies are subjected to a trial-by-trial analysis, the aim of which is to infer the processes that give rise to strategy change (Siegler and Crowley, 1991). While the focus on microgenetic methods hints at the multidimensional nature of individual differences in the acquisition of a specific ability, it has had relatively little to say about (1) the significance of different acquisition pathways, (2) the cognitive indices associated with different pathways, or (3) whether the same indices are relevant at different change points.

In the present study, we investigated changes in SDA problem solving speed variability (PRT) patterns longitudinally. The rationale for focusing on problem solving speed variability is we have found a close association between strategy-use and problem solving speed (Canobi et al., 1998, 2002; Paul and Reeve, 2016; Major et al., 2017). For example, a *count all* strategy, where each addend is individually enumerated, takes more time to execute and is more error prone than a *retrieval* strategy where answers are retrieved from memory (i.e., the answer is known and does not require computation). And, we have found a strategy-speed correlation independent of whether SDA problem was solved correctly or not (Paul and Reeve, 2016). We argue that the time taken for an individual to answer to a SDA problem is a defensible proxy for SDA strategy use (Major et al., 2017). Moreover, we can analyze the variability in SDA PRTs after individuals have learned to solve problems correctly.

Analyzing the variability in the speed with which individuals react to an event or solve problems has a long history in research on the neurophysiological basis of individual differences (Jensen, 1992). Indeed, it was pointed out 50 years ago that inter-event variability in RTs is not necessarily a measurement error in the narrow sense, but maybe a robust phenomenon in which there are reliable individual differences in RT patterns (Berkson and Baumeister, 1967). Recent research examining the RT patterns of children with ADHD, for example, shows they tend to have atypical RT patterns on attention tasks (Lewis et al., 2017). However, as far as we are aware, no research has investigated the significance of different RT patterns in SDA problem solving.

Cognitive Factors That May Affect SDA Strategies

A number of studies have investigated the association between cognitive factors (e.g., IQ, working memory), SDA strategy-use and problem solving ability (e.g., Paul and Reeve, 2016). Interpreting the importance of age-related factors responsible for general SDA abilities can be problematic since many abilities are correlated with age (Reeve et al., 2015). Furthermore, correlations tend to be modest, suggesting significant within-age variability in the factors affecting math abilities (Dowker, 2005). Nevertheless, associations have been found between SDA problem solving abilities, and some cognitive competencies (i.e., IQ, working memory) as well as core number abilities (dot enumeration, magnitude comparison) (Paul and Reeve, 2016). Increases in working-memory span (WM), for example, are associated with SDA problem solving accuracy (Raghubar et al., 2010). And poor WM capacity is thought to affect SDA strategies (e.g., by affecting the ability to monitor counting: see Dowker, 2005), and good WM capacity is associated with sophisticated SDA strategies (Geary, 2011; Geary et al., 2012). However, the association between the form of WM and math ability changes with age. In the young, math abilities tend to be correlated with visuospatial working memory (VSWM); and in older children verbal WM is more associated with math ability (De Smedt et al., 2009; Ashkenazi et al., 2013; van der Ven et al., 2013). This finding is consistent with the claim that visuospatial reasoning abilities are critical for early math (Gelman and Butterworth, 2005; Siegler and Mu, 2008; Dehaene and Brannon, 2011; Reeve et al., 2015).

In some studies non-verbal intelligence (NVIQ) is related to math abilities (Szűcs et al., 2014; however, see Reeve et al., 2012). In a longitudinal study Geary et al. (2017) reported that NVIQ was a stable predictor of children's math achievement (see also Van de Weijer-Bergsma et al., 2015; Lee and Bull, 2016; Tolar et al., 2016). One explanation for this association is NVIQ, in part, requires visuo-spatial abilities which are thought to be necessary for early math problem-solving ability (Szűcs et al., 2014). The question of the kinds of visuo-spatial skills that support different kinds of early math abilities is yet to be resolved, however.

Core number abilities are claimed to support early math development (Butterworth, 2010). The ability to rapidly and precisely enumerate small sets, for example, predicts concurrent and future math achievement (Reeve et al., 2012; Sasanguie et al., 2013; Bartelet et al., 2014; Gray and Reeve, 2014; Major et al., 2017). Dot enumeration tasks assess at least two components: a subitizing and a counting component. Subitizing is assessed by evaluating the way small sets ($n < 4$) are enumerated, which is usually accurately, rapidly and without error; counting is evaluated by assessing the way larger sets ($n > 4$) are enumerated, which usually more slowly and prone to counting errors (Schleifer and Landerl, 2011).

Reeve et al. (2012) identified three distinct dot enumeration profiles in 5-year-olds and showed profile membership remained stable over the primary school years. The three profiles differed in subitizing range, subitizing slope and intercept, but not counting slopes. Moreover, the profiles were associated with differences in math problems solving abilities. A similar pattern of findings has been observed in preschool children

(Gray and Reeve, 2014, 2016). We suggest that children with limited subitizing abilities may lack the ability to readily extract pattern or grouping information from small sets of dots (Butterworth, 2003; Ashkenazi et al., 2013). Why might this be important for numerical cognition? The ability to “know” the number “2” or “3” can be represented by a collection of two or three dots respectively, without counting individual dots, is an index of set knowledge (Butterworth, 2010); and set manipulation represents an important aspect of the development of numerical cognition (Gallistel and Gelman, 1992). The degree to which set knowledge changes in childhood is yet to be specified, however.

In recent research, Major et al. (2017) showed that dot enumeration profiles, in conjunction with performance on a standardized math test (the TEMA), assessed at school entry, predicted children's SDA problem solving speed longitudinally. However, the Major et al. findings were based on a general longitudinal path analytic model and their findings are silent about the possibility of different SDA PRT pathways, which is the focus of the current research.

The Current Research

The current research examined changes in children's SDA problem-solving response time variability (PRT) four times over 6 years (at 6, 7, 9, and 10 years) to determine whether it is possible to identify separate SDA PRT trajectories across time. Insofar as different speed trajectories could be identified, we investigated the degree to which different cognitive indices (i.e., VSWM assessed at 7 years, WM assessed at 9 years, speed naming numbers/letters, non-verbal IQ, and dot enumeration RTs in the subitizing range assessed at 9 years) were associated with different SDA PRT pathways; and the degree to which different SDA PRT pathways predicted performance on a double-digit mental addition (DDA assessed at 10 years) accounting for other cognitive abilities.

We included the VSWM and WM measures because math abilities tend to be correlated with VSWM in young children and with verbal WM in older children (Ashkenazi et al., 2013). We included the DDA task because on face-value it is a conceptually more complex version of the SDA task (see Major et al., 2017; Lemaire and Brun, 2018). Of interest is the degree to which different math acquisition pathways (i.e., variability in single digit addition problem speed over time) are associated with a common outcome. We included the naming numbers/letters speed task to assess for the possibility that findings reflect the speed with which information, particularly numerical information, is retrieved from memory.

We included the dot enumeration measure since previous research had shown that differences in responding to 1–3 dots is associated with math abilities at school entry and over the long term (Reeve et al., 2012; Major et al., 2017).

To identify different possible SDA speed variability trajectories over time, we used latent profile analysis (LPA) based on each individual child's mean variability in SDA PRTs at each of the four SDA assessment times. In LPA individuals are assigned to one of a number of subgroups or profiles that share common data patterns (Van Der Maas and Straatemeier, 2008). (This form of analysis has been used to characterize changes in the relationship between SDA strategy over time and VSWM—see Geary et al., 2009.)

Given the analytic focus of our research is change in the variability in SDA PRTs over time (rather than SDA strategy-use or problem-solving success), it would seem a priori reasonable to expect at least three PRT profiles to emerge from LPA: (1) a typical pathway in which mean SDA PRT variability diminishes over time; (2) a delayed SDA PRT pathway in which PRTs variability is high initially, but diminishes over time; and (3) a deficit pathway in which PRT variability remains relatively high over time. We acknowledge other profiles may emerge from LPA; however, we cannot anticipate what these might be a priori.

Insofar as SDA PRT variability pathways reflect different math specific (dot enumeration—subitizing ability, speed naming numbers) and/or general cognitive abilities (VSWM, WM, speed naming letters and NVIQ), we test several working hypotheses. Specifically, we expected children assigned to a delayed SDA PRT pathway would differ from children assigned to a typical pathway in their general cognitive capacities, but not their math specific ability (subitizing ability). Given the SDA PRTs of children in the delayed profile approach that of children in the typical profile over time, the delay is likely attributable to differences in general cognitive abilities. We expected children assigned to a deficit SDA PRT profile would differ from children in the typical and delayed pathway in their subitizing ability, and possibly their general cognitive abilities. This hypothesis is based on previous research which shows children with a math deficit also have poor subitizing abilities, but not necessarily general cognitive difficulties (Reeve et al., 2012; however see Gray and Reeve, 2016).

Insofar as different SDA PRT variability pathways reflect different arithmetic abilities, children assigned to the typical profile would be expected to perform better (would show less variability in response time and be more accurate) than those assigned to the delay profile who, in turn, would perform better than children assigned to a deficit profile on the double-digit mental addition task (DDA).

MATERIALS AND METHODS

Participants

One hundred-sixty-four children ($M = 72.59$ months, $SD = 4.58$ months at the beginning of the study), comprising 65 girls ($M = 71.52$ months, $SD = 4.47$ months) and 99 boys ($M = 73.29$ months, $SD = 4.54$ months), attending schools in middle-class suburbs of a large Australian city, participated in the study. All children spoke fluent English, had normal or corrected to normal vision and had no known learning disabilities (according to school personnel). The data reported herein were collected on four different occasions, namely, when children were 6, 7, 9, and 10 years of age. The children were part of a larger study investigating the development of math ability in preadolescent children across the primary/elementary school years (see Reeve et al., 2012 for details—note, only children who completed all assessments were included in the present study). At Time 2 children were 7-years-old ($M = 85.59$ months, $SD = 4.08$ months), at Time 3 children were 10-years-old ($M = 122.85$ months, $SD = 4.26$ months), and at Time 4 children were 11-years-old ($M = 129.49$ months, $SD = 4.55$ months).

The study was conducted in compliance with the requirements of the authors' University's Human Ethics Committee and the agreement of participating schools. Parents provided written consent allowing their child to participate in the project.

Materials and Procedure

Single-Digit Addition (Completed on All Four Occasions)

Twelve SDA problems were presented at each time point (see **Table 1**). Each pair of digits was presented in both orders (i.e., $2 + 5$ and $5 + 2$) to counterbalance and allow for the possibility to solve problems using a “min-counting” strategy (e.g., begin the count sequence from the largest addend to minimize the counting distance, irrespective of the fact that problems are read from left to right: see Paul and Reeve, 2016). Before beginning the task, children completed practice trials to familiarize them with the requirement to solve problems as quickly and as accurately as possible. Problems were presented in a random order. Problems appeared in the center of a 15" laptop screen in the form of $a + b = .$ Problem-solving accuracy and response times were recorded. The Chronbach's alphas, and associated 95% confidence interval for each SDA time measure, were—Time 1: 0.88 (0.85, 0.90); Time 2: 0.90 (0.87, 0.92); Time 3: 0.89 (0.86, 0.91); Time 4: 0.89 (0.87, 0.92).

Double-Digit Addition (Completed at 10 Years)

Twenty-four pairs of double-digit addend problems were presented (e.g., $28 + 19$), in which the sum of the addends was less than 100 (see **Table 2**). Problem-solving accuracy and response times were recorded (Cronbach's alpha = 0.95; 95% CI = 0.94 – 0.96.)

Forward Corsi Span (Completed at 6 Years)

The Corsi Blocks task (Milner, 1971) assessed visuo-spatial working memory, and was administered and scored following Kessels et al. (2000) procedure. An interviewer taps a sequence of blocks that attempts to repeat: beginning with two blocks, increasing by one block following each correct reproduction, up to a maximum of nine blocks. Testing concluded after two failed

TABLE 1 | Single-digit addition problem set repeated across Time 1 – Time 4.

Pair	Left addend	Right addend	Total sum	Numerical distance
Pair 1	2	4	6	2
Pair 2	4	2	6	2
Pair 3	2	5	7	3
Pair 4	5	2	7	3
Pair 5	3	5	8	2
Pair 6	5	3	8	2
Pair 7	2	6	8	4
Pair 8	6	2	8	4
Pair 9	3	6	9	3
Pair 10	6	3	9	3
Pair 11	2	7	9	5
Pair 12	7	2	9	5

TABLE 2 | Double-digit addition problems.

23 + 16	24 + 18	16 + 27
52 + 34	29 + 53	47 + 38
46 + 37	25 + 12	37 + 46
12 + 25	46 + 53	42 + 35
18 + 24	38 + 47	53 + 29
34 + 52	19 + 28	15 + 31
31 + 15	16 + 23	27 + 16
53 + 46	28 + 19	35 + 42

trials. The longest correct block tap sequence is the VSWM span. Reliability was $\alpha = 0.70$.

Backward Digit Span (Completed at 7 Years)

The backward version of the WISC-R Digit Span test was administered and scored as per the WISC-R Manual (Wechsler, 1986). This measure has been used to index WM capacity for verbal information (Geary et al., 2012). Reliability was $\alpha = 0.63$.

Naming Numbers Naming Letters (Completed at 9 Years)

In the naming numbers and naming letters tasks, the numbers 1–9 and the letters A–J (excluding the letter I because of its similarity to the number 1), respectively, were used. The two tasks comprised 36 trials, four each for the nine stimuli. The stimuli for both tasks, all of which were approximately 2 cm high on screen, were presented in one of four fixed random orders; the only constraint was that each stimulus should be different to the immediately preceding stimulus. Presentation order of the naming numbers and naming letters tasks was counterbalanced. (Cronbach's alphas: Naming Numbers = 0.96; 95% CI = 0.96 – 0.9; Naming Letters = 0.99; 95% CI = 0.99 – 0.99.)

Raven's Colored Progressive Matrices (RCPM) (Completed at 9 Years)

The RCPM is a measure of non-verbal IQ suitable for young children. It was included to assess the association between SDA processing speed and intelligence (Luwel et al., 2013). RCPM was administered following manual instructions (Raven et al., 1986), and scored using age norms (Raven et al., 1998). Research show good inter-item consistency and split-half reliability in a sample of Australian children (Cotton et al., 2005). The reliability estimate for the current sample was good ($\alpha = 0.82$).

Dot Enumeration (Completed at 10 Years)

Dot arrays comprising one to nine black dots (0.2 cm diameter) were presented on a white background. Dots were randomly positioned within a 15 cm × 11 cm grid and were no less than 2 cm apart (to reduce perceptual grouping cues). Each dot numerosity was presented eight times ($n = 72$ trials overall). Children were instructed to report as quickly and accurately as possible the number of dots in the array. Response accuracy and RTs were recorded. Here, only responses to dot arrays in the subitizing range (1–3 dots) were included in the analysis (24 trials). Previous research has shown differences in responding to 1–3 dots (i.e., differences in RTs, slope and intercept of the

subitizing range) is associated with math abilities at school entry and over the long term (Reeve et al., 2012; Major et al., 2017). However, the speed enumerating dots in the counting range (5–8 dots) was not associated with math ability (Reeve et al., 2012). It is worth noting that Anobile et al. (2016) showed that numerosity, but not texture-density, discrimination correlates with math ability in children. (Cronbach's alpha = 0.83; 95% CI = 0.79 – 0.86.)

Rationale for Measures

We calculated a measure of SDA problem-solving RT variability (SDA_{var}) for each child (i) by subtracting their average RT (μ_i) at each time point from each of the twelve SDA problems (q_j), and then taking the sum of the absolute values of these deviations ($|\mu - q|$):

$$SDA_{var} = \sum_{i=1}^n |q_i - \mu_i|$$

The same procedure was used to create RT variability measures for the naming numbers RT_{var} (nine trials), naming letters RT_{var} (nine trials), dot enumeration (DE_{var} , 24 trials), and double-digit addition (DDA_{var} , 24 problems) tasks. For dot enumeration, only responses to dot arrays in the subitizing range (1–3 dots) were included in this analysis (24 trials). Previous research has shown differences in responding to 1–3 dots (i.e., differences in RTs, slope and intercept of the subitizing range) is associated with math abilities at school entry and over the long term (Reeve et al., 2012; Major et al., 2017).

Corsi span (VSWM) scores represent the average of two trials of the forward version of the task (see Kessels et al., 2000). Digit span scores were measured as the sum of the forward and backward versions of the WISC-R test. Raven's (NVIQ) raw scores are used in analyses since scaled percentile scores were at ceiling level and non-normally distributed.

Analytic Approach

We used MPlus (Muthén and Muthén, 1998–2013) latent class/profile analysis to identify SDA problem-solving speed profiles. (It should be noted that we did not examine SDA problem solving success – most children were performing at ceiling on the second test occasions.) We estimated three LPA models with an increasing number of profiles based on expected patterns of change in SDA problem-solving variability over time: (1) a two profile solution would differentiate a typical pathway (e.g., decreased variability over time) from a deficit pathway (e.g., minimal decrease in variability over time); (2) a three profile solution would differentiate a typical pathway, a delayed pathway (e.g., slower decrease in variability over time compared to typical performance) and a deficit pathway; and (3) a four profile solution was expected to identify a typical pathway, a delayed pathway and a deficit pathway, while also allowing for the possibility of another different pathway (e.g., irregular shifts in variability over time).

Once profiles were identified, children were allocated to the profile with the highest probability of membership. To further distinguish between these pathways, One-way ANOVAs were conducted to characterize differences in measures of

cognitive ability (i.e., VSWM, naming numbers RT variability, naming letter RT variability, digit span, NVIQ, and subitizing RT variability) between the profiles. One-way ANOVAs were also conducted to determine whether profiles were associated with double-digit addition problem-solving accuracy and response time variability. Regression analyses were conducted to determine the independent contribution of the profiles and cognitive abilities in predicting double-digit addition problem-solving accuracy and response time variability.

RESULTS

Descriptive Statistics

Bivariate correlations and means (standard deviations) for measures are reported in **Table 3**. Of note, SDA PRT variability (SDA_{var} 6, 7, 9, and 10 years) showed an average decrease in PRT over time; however, there was significant variation in means over time, suggesting different patterns of variance may be embedded within the overall variance. We used LPA to investigate this possibility.

SDA Problem Solving Speed Profiles

Latent profile models with two to four profiles were compared in terms of different goodness-of-fit indices to determine the best-fitting solution to the data. **Table 4** shows all relative fit statistics (AIC, BIC and aBIC) improved for models with an increasing number of profiles and entropy values were high (≥ 0.8 , suggesting good separation of profiles; Clark and Muthén, 2009). While the four-profile solution provided better fit than the three-profile solution (i.e., significant bootstrap likelihood-ratio test scores; see **Table 4**), examination of the four profiles revealed two profiles were similar—both profiles showed patterns of delayed decrease in variability over time, which were not meaningfully different from each other. The three-profile solution characterized more distinct patterns of change in variability over time, and were more consistent with typical, delayed and deficit pathways. Since the three-profile model was a more parsimonious description of the data and was less likely to lead to over-fitting our sample than the four-profile model, the three-profile model was selected for further examination (see **Supplementary Material**).

The three profiles differed in mean RT and SDA variability measures over time (see **Figure 1**). The first pathway (Typical pathway, $n = 71$, 43.3%) showed a decrease in SDA problem-solving speed variability over time, and exhibited minimal variability at Times 3 and 4. The second pathway (Delayed pathway, $n = 78$, 47.6%) showed a similar decrease in RT variability over time; however, the variability was still decreasing at Times 3 and 4. The third pathway (Deficit pathway, $n = 15$, 9.1%) showed a decrease in RT over time but SDA variability remained high.

Analysis of Cognitive Abilities Across SDA PRT Profiles

One-way ANOVAs were conducted to determine whether the measures of cognitive ability differed across the three

profiles. Bias-corrected and accelerated bootstrap estimates (95% confidence, 1000 draws) are reported to account for unequal variance between profiles, and Welch correction for robust test of equality of means was applied when necessary. The profiles differed significantly in terms of subitizing RT_{var} [$F_{WELCH}(2, 34.93) = 8.48$, $p = 0.001$, Levine = 17.37, $p < 0.001$]. The Typical pathway had significantly lower subitizing RT_{var} compared to the Delayed ($p = 0.020$) and Deficit ($p = 0.011$) pathways, while the Delayed pathway had significantly lower subitizing RT_{var} compared to the Deficit ($p = 0.046$) pathway. The pathways did not significantly differ in terms of VSWM span [$F(2,161) = 0.89$, $p = 0.413$], naming numbers RT_{var} [$F(2,161) = 0.62$, $p = 0.540$] or naming letters RT_{var} [$F(2,161) = 0.59$, $p = 0.553$], digit span [$F(2,161) = 1.96$, $p = 0.144$] or NVIQ [$F(2,161) = 1.49$, $p = 0.228$].

Association Between Variability Pathways and Double Digit Addition Ability

A one-way ANOVA (bias-corrected and accelerated bootstrap estimates) compared the double-digit problem-solving accuracy across the three profiles. Double-digit accuracy differed significantly between the profiles [$F_{WELCH}(2,35.22) = 13.12$, $p < 0.001$, Levine = 19.97, $p < 0.001$]. Post hoc comparisons (corrected for unequal variances, Games-Howell) showed the *Typical Pathway* ($M = 0.92$, $SD = 0.09$) had significantly higher double-digit problem-solving accuracy than the *Delayed Pathway* ($p < 0.001$) and *Deficit Pathway* ($p = 0.016$); the *Deficit Pathway* ($M = 0.75$, $SD = 0.20$) had the lowest double-digit problem-solving accuracy, but was not significantly different from the *Delayed Pathway* ($M = 0.83$, $SD = 0.16$).

A separate one-way ANOVA compared response time variability between profiles, which showed double-digit response time variability differed significantly across profiles [$F_{WELCH}(2,35.65) = 25.10$, $p < 0.001$, Levine = 9.67, $p < 0.001$]. Post-hoc comparisons showed the *Typical Pathway* ($M = 53.55$, $SD = 27.12$) had significantly lower double-digit response time variability than both *Delayed Pathway* ($p = 0.001$) and *Deficit Pathway* ($p < 0.001$); the *Delayed Pathway* ($M = 85.20$, $SD = 38.93$) had significantly lower double-digit response time variability than the *Deficit Pathway* ($p = 0.037$); and the *Deficit Pathway* ($M = 126.96$, $SD = 56.89$) had the highest double-digit response time variability.

The cognitive abilities and pathway membership (dummy coded relative to the *Deficit Pathway*) were entered into separate linear regression analyses to determine the degree to which they predicted double-digit addition problem-solving accuracy (Model 1, **Table 5**) and response time variability (Model 2, **Table 6**). (Note, we report separate analyses that included/exclude the pathways for clarify sake.) Overall, only the subitizing measure significantly predicted DDA accuracy [Model 1a (with cognitive abilities): $F(6,157) = 3.19$, $p = 0.006$; Model 1b (with variability profiles): $F(8,155) = 4.42$, $p < 0.001$] and response time variability [Model 2a (with cognitive abilities): $F(6,157) = 7.16$, $p < 0.001$]; Model 2b (with variability profiles): $F(8,155) = 10.55$, $p < 0.001$].

TABLE 3 | Bivariate correlations and means (standard deviations) for all measures.

	1	2	3	4	5	6	7	8	9	10	11	12
(1) SDA _{var} 6 years	–											
(2) SDA _{var} 7 years	0.17*	–										
(3) SDA _{var} 9 years	0.09	0.33***	–									
(4) SDA _{var} 10 years	0.09	0.36***	0.65***	–								
(5) VSWM	–0.09	–0.07	–0.06	–0.18*	–							
(6) Naming numbers RT _{var}	–0.04	0.01	0.07	–0.03	–0.05	–						
(7) Naming letters RT _{var}	–0.04	0.05	0.09	–0.01	–0.08	0.84**	–					
(8) Digit span	–0.14	–0.12	–0.03	–0.03	0.05	0.09	0.08	–				
(9) NVIQ	0.08	–0.13	–0.06	–0.04	0.23**	–0.11	–0.15	0.28**	–			
(10) DE _{var}	–0.08	0.29**	0.45**	0.55**	–0.14	0.05	0.09	0.00	–0.09	–		
(11) DDA _{acc}	–0.27**	–0.26**	–0.34**	–0.27**	0.04	–0.03	–0.07	0.12	0.01	–0.29**	–	
(12) DDA _{var}	0.25**	0.42**	0.40**	0.58**	–0.05	–0.03	–0.02	–0.15	–0.00	0.43**	–0.52**	–
Mean	39.99	23.92	8.93	7.89	3.66	3.38	4.44	9.13	31.64	4.81	0.86	75.32
SD	27.47	16.79	8.29	6.34	0.65	0.83	1.31	2.19	3.51	2.22	0.15	42.53

SDA_{var} is the sum of trial-by-trial variability in single-digit addition problem-solving (12 trials); VSWM is the average block length across two forward trials; Naming numbers_{var} and Naming letters_{var} is the sum of trial-by-trial variability (nine trials); Digit span is the sum of sequence length across forward and backward trials; NVIQ is the raw Raven's score; DE_{var} is the sum of trial-by-trial variability (24 trials); DDA_{acc} is the accuracy of double-digit addition problem-solving; and DDA_{acc} is the sum of trial-by-trial variability in double-digit addition problem-solving (24 trials). *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

TABLE 4 | Latent profile analysis goodness-of-fit indices.

Profiles	Parameters	LL	AIC	BIC	aBIC	BLRT	Entropy
2	17	–2425.42	4884.83	4937.53	4883.71	<0.001	0.86
3	26	–2373.91	4799.83	4880.42	4798.11	<0.001	0.83
4	35	–2342.39	4754.78	4863.28	4752.47	<0.001	0.85

LL, Log-likelihood; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; aBIC, Adjusted Bayesian Information Criterion; BLRT, Bootstrap Likelihood-Ratio Test (100 draws).

DISCUSSION

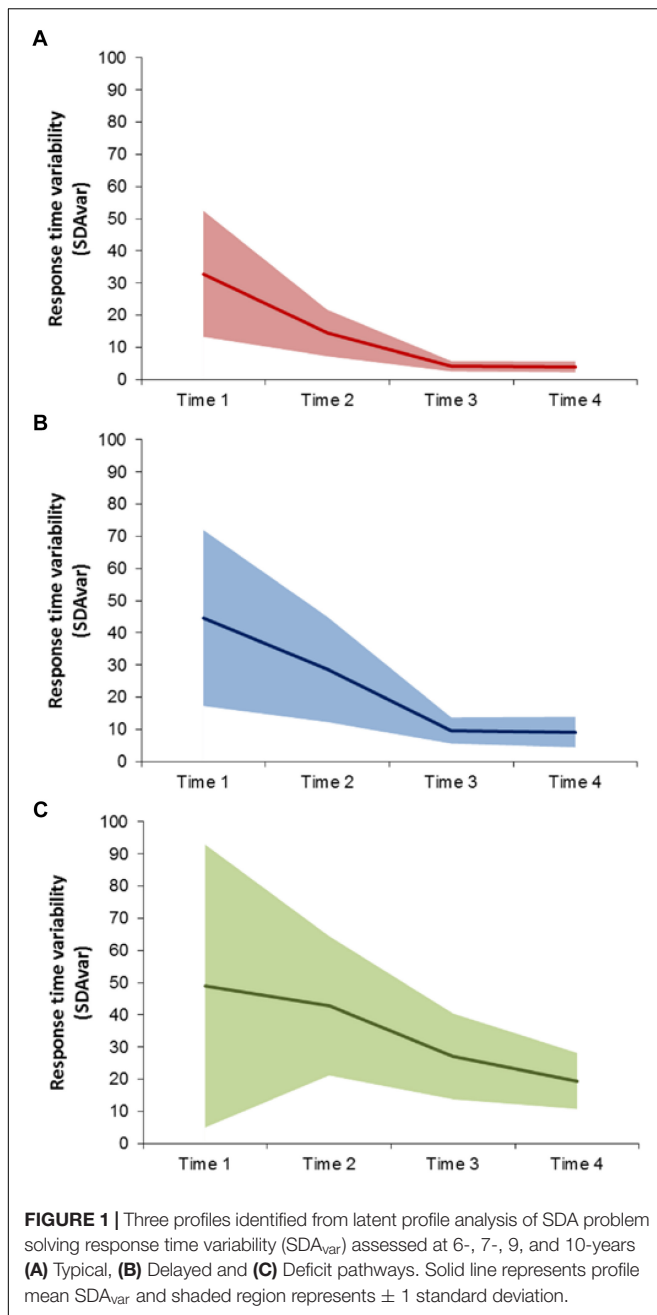
The study investigated whether different patterns of change in SDA PRT trajectories in primary/elementary aged children could be identified over a 6 years period, and the degree to which these patterns reflect typical, delayed or deficit math acquisition pathways. It also assessed the degree to which different SDA PRT change pathways were associated with differences in VSWM, WM, NVIQ, digit naming and subitizing speed, as well as the degree to which the different SDA PRT pathways predicted double digit mental addition problem solving speed and accuracy.

Four findings are of note. First, three distinctly different SDA PRT pathways were identified. In one, labeled a typical acquisition pathway, mean SDA PRT was relatively fast, with relatively little PRT variability. In the second, labeled a delayed pathway, both SDA PRT means and variability were high initially, but diminished over time. In the third pathway, labeled a deficit pathway, SDA PRT mean and variability remained relatively high over the 6 years assessment period. As noted earlier, nearly all children were able to solve SDA problems correctly. Second, with one exception, the three SDA PRT pathways differed in the subitizing variability measure only, and no other cognitive measures. The exception was WM was associated with DDA problem solving success. Third, the subitizing variability measure remained associated with both the DDA success and

variability measures, after the pathway factor had been included in regression equations. Fourth, the typical pathway contributed to the equation predicting DDA variability over and above the deficit pathway; and the delayed pathway over and above the deficit pathway. And, the typical pathway contributed to the equation predicting DDA problem solving success over and above the deficit pathway; however, the delayed pathway did not contribute to the prediction equation over and above the deficit pathway.

The pattern of findings support the claims that (1) speed variability signatures are associated with math problem solving ability, even when problems are solved correctly, (2) with the exception of subitizing speed signatures, standard cognitive indices appear unrelated to SDA speed variability indices; and (3) variability in dot enumeration speed signatures within the subitizing range predicts math ability (at least, double-digit mental addition ability). The question remains, why are dot enumeration speed variability signatures specifically, and problem solving variability signatures generally, a predictor of individual difference in math abilities? One answer to this question lies in understanding the reason(s) for differences in dot enumeration subitizing ability.

In a series of studies, we have shown that dot enumeration abilities, and subitizing ability in particular, are associated with children's math abilities (Reeve et al., 2012;



Gray and Reeve, 2014, 2016; Major et al., 2017). In large measure, these studies were motivated by a desire to better understand the reasons for individual differences in Butterworth's dot enumeration task (see Butterworth's, 2003, "Dyscalculia Screener"). Reeve et al. (2012) showed that individual differences in children's subitizing abilities (indexed by the subitizing range, slope and intercept) assessed at school entry predicted math performance across the primary/elementary school years. While these subitizing indices "improved" across time, children's performance changed at a relative rate compared to each other (i.e., rank order correlations remain stable). Moreover, Major et al. (2017) showed that subitizing abilities assessed at school

entry was as good a predictor of school math performances as performance on a standardized math test (The Test of Early Mathematics Ability) in the short term, and a much better predictor in the long term. Furthermore, Gray and Reeve (2014, 2016) showed that pre-schooler's dot enumeration abilities also predict their emerging math abilities. Other researchers have also found a relationship between subitizing dot enumeration and poor math abilities (Desoete et al., 2009; Reigosa-Crespo et al., 2012; Landerl, 2013).

These findings indirectly emphasize the importance of variability in subitizing speed as a predictor of math ability, but not the reason(s) for its importance. We suggest that poor subitizing abilities reflect a lack an ability to readily extract pattern or grouping information from small sets of dots (Butterworth, 2003; Ashkenazi et al., 2013). Why might this be important for numerical cognition? The ability to "know" the number "2" or "3" can be represented by a collection of two or three dots respectively, without counting individual dots, is arguably a fundamental index of set knowledge (Butterworth, 2010). In the absence of "automatic" set extraction ability, individuals would need to count individual dots. Indeed, set manipulation ability is argued to be an important ability in the development of numerical cognition (Gallistel and Gelman, 1992). We suggest the three speed profiles identified herein reflect different levels of set extraction ability. In the absence of set knowledge, numerical reasoning is likely to be difficult, as is evident in individuals with developmental dyscalculia, who appear to lack the ability to extract information from small sets of dots at a glance (Butterworth, 2010).

The number of children assigned to the deficit profile in the current analysis (8.5% of the sample) is similar to the number of children thought to possess dyscalculia in general population (see Butterworth, 2010). Insofar as the variability in the speed with which small arrays of dots are enumerated is an index of set ability, it is reasonable to ask whether it is a general cognitive or a number specific constraint. It has long been claimed that processing speed is a proxy measure of intelligence (Coyle et al., 2011; however, see Cepeda et al., 2013). Caution should be exercised, however, in arguing for a general processing speed hypothesis on the basis of our findings for two reasons. First, the focus of our research was variability in the speed with which children solve number problems, rather than speed *per se*. Second, while SDA PRT variability and subitizing RT variability independently contributed to the equation predicting double-digit mental addition (success and response time variability), it is difficult to specify the reason(s) for this independence. The acquisition of math ability comprises different components, the importance of which likely varies with age (Dowker, 2005; Gray and Reeve, 2016). It is possible that effective set abilities in the young facilitate the emergence of other math skills, including SDA abilities.

It is worth noting that the variability in speed with which children named the numbers one to nine and the letter A to J was unrelated to other speed variability measures, which argued against the claim that speed variability is a general cognitive constraint, and rather supports the claim that it is number-specific constraint.

TABLE 5 | Model 1: Linear regressions predicting DDA PRT success.

	<i>Beta</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Model 1a – Cognitive Abilities					
VSWM span	0.01	0.03	0.05	0.32	0.793
Naming numbers RT	−0.01	0.02	−0.10	−0.71	0.590
Naming letters RT	0.00	0.02	0.00	0.01	0.992
Digit span	0.01	0.01	0.15	1.84	0.050
NVIQ	−0.00	0.00	−0.07	−0.84	0.480
Subitizing RT	−0.02	0.01	−0.29	−3.83	0.001**
Model 1b – Variability Profiles					
VSWM span	0.01	0.02	0.04	0.31	0.785
Naming numbers RT	−0.01	0.02	−0.12	−0.88	0.458
Naming letters RT	−0.00	0.02	−0.02	−0.20	0.857
Digit span	0.01	0.01	0.11	1.40	0.130
NVIQ	−0.00	0.00	−0.08	−0.96	0.447
Subitizing RT	−0.01	0.01	−0.19	−2.37	0.023*
Typical Pathway ^a	0.12	0.04	0.39	2.63	0.036*
Delayed Pathway ^a	0.03	0.04	0.11	0.81	0.550

^aDummy-coded relative to the Deficit pathway. *P*-values are bias-corrected accelerated bootstrap estimates (1000 samples). ***p* < 0.01, **p* < 0.05.

TABLE 6 | Model 2: Linear regression predicting DDA PRT variability.

	<i>Beta</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Model 2a – Cognitive abilities					
VSWM span	−5.21	6.70	−0.10	−0.78	0.506
Naming numbers RT	2.90	4.29	0.09	0.68	0.573
Naming letters RT	0.47	4.86	0.01	0.10	0.913
Digit span	−3.38	1.45	−0.17	−2.33	0.053
NVIQ	1.01	0.94	0.08	1.07	0.388
Subitizing RT	8.27	1.37	0.43	6.02	0.001**
Model 2b – Variability Profiles					
VSWM span	−5.30	6.13	−0.14	−0.87	0.452
Naming numbers RT	3.71	3.92	0.12	0.95	0.433
Naming letters RT	1.39	4.45	0.02	0.31	0.740
Digit span	−2.29	1.34	−0.12	−1.71	0.153
NVIQ	1.22	0.86	0.10	1.41	0.245
Subitizing RT	5.22	1.39	0.27	3.75	0.003**
Typical pathway ^a	−54.56	11.25	−0.64	−4.85	0.001**
Delayed pathway ^a	−27.38	10.70	−0.32	−2.56	0.062

^aDummy-coded relative to the Deficit pathway. *P*-values are bias-corrected accelerated bootstrap estimates (1000 samples). ***p* < 0.01.

Limitations of Research

In the present study we examined the variability in SDA problem solving speed. On the basis of our previous research, we are reasonably confident problem solving speed reflects SDA strategy-use—immature SDA strategies take longer to execute than more mature strategies (see Canobi et al., 1998, 2002; Paul and Reeve, 2016). Nevertheless, we did not examine the mix of SDA problem solving strategies, or how this mix changes in the typical, delayed and deficit groups over time. It is possible that the speed variability measure may obscure other indices (e.g., variability in speed taken to execute the same SDA strategy over time).

While we have argued for a distinction between a typical, delayed, and deficit math pathway, it is important not to overstate the robustness of this argument for two reasons. First, we have focused on a relatively narrow range of computation abilities (SDA and DDA) over a relatively short time. It is possible, with time, the performance of children in the deficit and delayed groups would approach the performance of children in the typical pathway group. Second, although we focused on mental addition in the pre-adolescent years because of its importance in math curricula, we recognize the pattern of findings may differ for other math competencies (e.g., subtraction, multiplication, division).

CONCLUSION

We have argued that the variability in the speed with which children enumerate one to three dots is an index of the ability to rapidly extract set knowledge, which, in turn, is a key ingredient in the acquisition of preadolescent children's math ability. However, the degree to which set knowledge changes in childhood is yet to be specified precisely, or the degree to which it is supported by other cognitive functions (e.g., attention abilities). Nevertheless, we suggest our findings have diagnostic and intervention implications. Given the variability in dot enumeration RTs is a diagnostic measure of math ability, it is a relatively easy measure to collect and interpret.

ETHICS STATEMENT

This study was conducted in accordance with the recommendations of the Human Ethics Committee of the University of Melbourne. Written informed consent was obtained from the parents of the children who participated in the study. All participants provide informed consent in accordance

with the Declaration of Helsinki. The interview protocols were approved by the Human Research Ethics Committee of the University of Melbourne and by participating schools.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

The research reported herein was supported by an Australian Research Award (DP0557199) to RR and BB.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01498/full#supplementary-material>

REFERENCES

- Anobile, G., Castaldi, E., Turi, M., Tinelli, F., and Burr, D. C. (2016). Numerosity but not texture-density discrimination correlates with math ability in children. *Dev. Psychol.* 52, 1206–1216. doi: 10.1037/dev0000155
- Ashkenazi, S., Rosenberg-Lee, M., Metcalfe, A. W. S., Swigart, A. G., and Menon, V. (2013). Visuo-spatial working memory is an important source of domain-general vulnerability in the development of arithmetic cognition. *Neuropsychologia* 51, 2305–2317. doi: 10.1016/j.neuropsychologia.2013.06.031
- Baroody, A. J. (2003). "The development of adaptive expertise and flexibility: the integration of conceptual and procedural knowledge," in *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise*, eds A. J. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates), 1–34.
- Baroody, A. J., and Dowker, A. (2003). *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise*. Mahwah, NJ: Erlbaum.
- Bartelet, D., Vaessen, A., Blomert, L., and Ansari, D. (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *J. Exp. Child Psychol.* 117, 12–28. doi: 10.1016/j.jecp.2013.08.010
- Berkson, G., and Baumeister, A. (1967). Reaction time variability of mental defectives and normals. *Am. J. Ment. Defic.* 72, 262–266.
- Butterworth, B. (2003). *Dyscalculia Screener*. London: NFER Nelson Publishing Company Ltd.
- Butterworth, B. (2005). The development of arithmetical abilities. *J. Child Psychol. Psychiatry* 1, 3–18. doi: 10.1111/j.1469-7610.2004.00374.x
- Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends Cogn. Sci.* 14, 534–541. doi: 10.1016/j.tics.2010.09.007
- Canobi, K. H., Reeve, R. A., and Pattison, P. E. (1998). The role of conceptual understanding in children's addition problem solving. *Dev. Psychol.* 34, 882–891. doi: 10.1037/0012-1649.34.5.882
- Canobi, K. H., Reeve, R. A., and Pattison, P. E. (2002). Young children's understanding of addition concepts. *Educ. Psychol.* 22, 513–532. doi: 10.1080/0144341022000023608
- Cepeda, N. J., Blackwell, K. A., and Munakata, Y. (2013). Speed isn't everything: complex processing speed measures mask individual differences and developmental changes in executive control. *Dev. Sci.* 16, 269–286. doi: 10.1111/desc.12024
- Clark, S., and Muthén, B. O. (2009). *Relating Latent Class Analysis Results to Variables not Included in the Analysis*. Available at: <http://www.statmodel.com/download/relatinglca.pdf>
- Cotton, S. M., Kiely, P. M., Crewther, D. P., Thomson, B., Laycock, R., and Crewther, S. G. (2005). A normative and reliability study for the Raven's coloured progressive matrices for primary school aged children from Victoria, Australia. *Pers. Individ. Dif.* 39, 647–659. doi: 10.1016/j.paid.2005.02.015
- Coyle, T. R., Pillow, D. R., Snyder, A. C., and Kochunov, P. (2011). Processing speed mediates the development of general intelligence (g) in adolescence. *Psychol. Sci.* 22, 1265–1269. doi: 10.1177/0956797611418243
- De Smedt, B., Verschaffel, L., and Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *J. Exp. Child Psychol.* 103, 469–479. doi: 10.1016/j.jecp.2009.01.010
- Dehaene, S., and Brannon, E. (2011). *Space, Time, and Number in the Brain: Searching for the Foundations of Mathematical Thought*. London: Elsevier.
- Desoete, A., Ceulemans, A., Roeyers, H., and Huybrechts, A. (2009). Subitizing or counting as possible screening variables for learning disabilities in mathematics education or learning? *Educ. Res. Rev.* 4, 55–66. doi: 10.1016/j.edurev.2008.11.003
- Dowker, A. (2005). Early identification and intervention for students with mathematics difficulties. *J. Learn. Disabil.* 38, 324–332. doi: 10.1177/00222194050380040801
- Gallistel, C., and Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition* 44, 43–74. doi: 10.1016/0010-0277(92)90050-R
- Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Dev. Psychol.* 47, 1539–1552. doi: 10.1037/a0025510
- Geary, D. C., Bailey, D. H., Littlefield, A., Wood, P., Hoard, M. K., and Nugent, L. (2009). First-grade predictors of mathematical learning disability: a latent class trajectory analysis. *Cogn. Dev.* 24, 411–429. doi: 10.1016/j.cogdev.2009.10.001
- Geary, D. C., and Hoard, M. K. (2005). "Learning disabilities in arithmetic and mathematics," in *Handbook of Mathematical Cognition*, ed. J. I. D. Campbell (Hove: Psychology Press), 253–267.
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., and Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Dev.* 78, 1343–1359. doi: 10.1111/j.1467-8624.2007.01069.x
- Geary, D. C., Hoard, M. K., Nugent, L., and Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low

- achievement: a five-year prospective study. *J. Educ. Psychol.* 104, 206–223. doi: 10.1037/a0025398
- Geary, D. C., Nicholas, A., Li, Y., and Sun, J. (2017). Developmental change in the influence of domain-general abilities and domain-specific knowledge on mathematics achievement: an eight-year longitudinal study. *J. Educ. Psychol.* 109, 680–693. doi: 10.1037/edu0000159
- Gelman, R., and Butterworth, B. (2005). Number and language: how are they related? *Trends Cogn. Sci.* 9, 6–10.
- Gray, S. A., and Reeve, R. A. (2014). Preschoolers' dot enumeration abilities are markers of their arithmetic competence. *PLoS One* 9:e94428. doi: 10.1371/journal.pone.0094428
- Gray, S. A., and Reeve, R. A. (2016). Number-specific and general cognitive markers of preschoolers' math ability profiles. *J. Exp. Child Psychol.* 147, 1–21. doi: 10.1016/j.jecp.2016.02.004
- Jensen, A. R. (1992). The importance of intraindividual in reaction time variation. *Pers. Individ. Dif.* 13, 869–881. doi: 10.1016/0191-8869(92)90004-9
- Jordan, J. A., Mulhern, G., and Wylie, J. (2009). Individual differences in trajectories of arithmetical development in typically achieving 5- to 7-year-olds. *J. Exp. Child Psychol.* 103, 455–468. doi: 10.1016/j.jecp.2009.01.011
- Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., and de Haan, E. H. F. (2000). The corsi block-tapping task: standardization and normative data. *Appl. Neuropsychol.* 7, 252–258. doi: 10.1207/S15324826AN0704_8
- Landerl, K. (2013). Development of numerical processing in children with typical and dyscalculic arithmetic skills—a longitudinal study. *Front. Psychol.* 4:459. doi: 10.3389/fpsyg.2013.00459
- Lee, K., and Bull, R. (2016). Developmental changes in working memory, updating, and math achievement. *J. Educ. Psychol.* 108, 869–882. doi: 10.1037/edu0000090
- Lemaire, P., and Brun, F. (2018). Age-related changes in children's strategies for solving two-digit addition problems. *J. Numer. Cogn.* 3, 582–597. doi: 10.1007/s00426-013-0501-0
- Lewis, F. C., Reeve, R. A., Kelly, S. P., and Johnson, K. A. (2017). Sustained attention to a predictable, unengaging Go/No-Go task shows ongoing development between 6 and 11 years. *Atten. Percept. Psychophys.* 79, 1726–1741. doi: 10.3758/s13414-017-1351-4
- Luwel, K., Foustana, A., Onghena, P., and Verschaffel, L. (2013). The role of verbal and performance intelligence in children's strategy selection and execution. *Learn. Individ. Dif.* 24, 134–138. doi: 10.1016/j.lindif.2013.01.010
- Major, C. S., Paul, J. M., and Reeve, R. A. (2017). TEMA and dot enumeration profiles predict mental addition problem solving speed longitudinally. *Front. Psychol.* 8:2263. doi: 10.3389/fpsyg.2017.02263
- Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *Br. Med. Bull.* 27, 272–277. doi: 10.1093/oxfordjournals.bmb.a070866
- Muthén, L. K., and Muthén, B. O. (1998–2013). *Mplus User's Guide*, 7th Edn. Los Angeles, CA: Muthén & Muthén.
- OECD. (2014). *Education at a Glance 2014: OECD Indicators*. Paris: OECD Publications.
- Paul, J. M., and Reeve, R. A. (2016). Relationship between single digit addition strategies and working memory reflects general reasoning sophistication. *Learn. Instr.* 42, 113–122. doi: 10.1016/j.learninstruc.2016.01.011
- Raghubar, K. P., Barnes, M. A., and Hecht, S. A. (2010). Working memory and mathematics: a review of developmental, individual difference, and cognitive approaches. *Learn. Individ. Dif.* 20, 110–122. doi: 10.1016/j.lindif.2009.10.005
- Raven, J., Raven, J. C., and Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales, Section 3: The Standard Progressive Matrices*. Oxford: Oxford Psychologists Press.
- Raven, J. C., Court, J. H., and Raven, J. (1986). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. London: Lewis.
- Reeve, R., Reynolds, F., Humberstone, J., and Butterworth, B. (2012). Stability and change in markers of core numerical competencies. *J. Exp. Psychol.* 141, 649–666. doi: 10.1037/a0027520
- Reeve, R. A., Paul, J. M., and Butterworth, B. (2015). Longitudinal changes in young children's 0-100 to 0-1000 number-line error signatures. *Front. Psychol.* 6:647. doi: 10.3389/fpsyg.2015.00647
- Reigosa-Crespo, V., Valdés-Sosa, M., Butterworth, B., Estévez, N., Rodríguez, M., Santos, E., et al. (2012). Basic numerical capacities and prevalence of developmental dyscalculia: the Havana survey. *Dev. Psychol.* 48, 123–135. doi: 10.1037/a0025356
- Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number – space mappings: what underlies mathematics achievement? *J. Exp. Child Psychol.* 114, 418–431. doi: 10.1016/j.jecp.2012.10.012
- Schleifer, P., and Landerl, K. (2011). Subitizing and counting in typical and atypical development. *Dev. Sci.* 14, 280–291. doi: 10.1111/j.1467-7687.2010.00976.x
- Siegler, R. S. (1996). *Emerging Minds: The Process of Change in Children's Thinking*. New York, NY: Oxford University Press.
- Siegler, R. S. (2000). The rebirth of children's learning. *Child Dev.* 71, 26–35. doi: 10.1111/1467-8624.00115
- Siegler, R. S. (2005). Children's learning. *Am. Psychol.* 60, 769–778. doi: 10.1037/0003-066X.60.8.769
- Siegler, R. S. (2006). "Microgenetic analyses of learning," in *Handbook of Child Psychology: Volume 2: Cognition, Perception, and Language*, eds W. Damon, R. M. Lerner, D. Kuhn, and R. S. Siegler (Hoboken, NJ: Wiley), 464–510.
- Siegler, R. S. (2007). Cognitive variability. *Dev. Sci.* 10, 104–109. doi: 10.1111/j.1467-7687.2007.00571.x
- Siegler, R. S. (2016). Continuity and change in the field of cognitive development and in the perspectives of one cognitive developmentalist. *Child Dev. Perspect.* 10, 128–133. doi: 10.1111/cdep.12173
- Siegler, R. S., and Crowley, K. (1991). The microgenetic method: a direct means for studying cognitive development. *Am. Psychol.* 46, 606–620. doi: 10.1037/0003-066X.46.6.606
- Siegler, R. S., and Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychol. Sci.* 19, 759–763. doi: 10.1111/j.1467-9280.2008.02153.x
- Szűcs, D., Devine, A., Soltesz, F., Nobes, A., and Gabriel, F. (2014). Cognitive components of a mathematical processing network in 9-year-old children. *Dev. Sci.* 17, 506–524. doi: 10.1111/desc.12144
- Tolar, T. D., Fuchs, L., Fletcher, J. M., Fuchs, D., and Hamlett, C. L. (2016). Cognitive profiles of mathematical problem solving learning disability for different definitions of disability. *J. Learn. Disabil.* 49, 240–256. doi: 10.1177/0022219414538520
- Torbeys, J., Arnaud, L., Lemaire, P., and Verschaffel, L. (2004). "Cognitive change as strategic change," in *Cognitive Developmental Change: Theories, Models and Measurement*, eds A. Demetriou and A. Raftopoulos (Cambridge: Cambridge University Press), 186–216.
- Van de Weijer-Bergsma, E., Kroesbergen, E. H., and Van Luit, J. E. H. (2015). Verbal and visual-spatial working memory and mathematical ability in different domains throughout primary school. *Mem. Cognit.* 43, 367–378. doi: 10.3758/s13421-014-0480-4
- Van Der Maas, H. L. J., and Straatemeier, M. (2008). How to detect cognitive strategies: commentary on "Differentiation and integration: guiding principles for analyzing cognitive change." *Dev. Sci.* 11, 449–453. doi: 10.1111/j.1467-7687.2008.00690.x
- van der Ven, S. H. G., van der Maas, H. L. J., Straatemeier, M., and Jansen, B. R. J. (2013). Visuospatial working memory and mathematical ability at different ages throughout primary school. *Learn. Individ. Dif.* 27, 182–192. doi: 10.1016/j.lindif.2013.09.003
- Wechsler, D. (1986). *Manual for the Wechsler Intelligence Scale for Children - Revised*. New York, NY: Harcourt Test Publishers.

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

Copyright © 2018 Reeve, Gray, Butterworth and Paul. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Training Can Increase Students' Choices for Written Solution Strategies and Performance in Solving Multi-Digit Division Problems

Marije F. Fagginger Auer^{1,2}, Marian Hickendorff^{3*} and Cornelis M. van Putten¹

¹ Methodology and Statistics, Institute of Psychology, Leiden University, Leiden, Netherlands, ² The Netherlands Association of Universities of Applied Sciences, The Hague, Netherlands, ³ Educational Science, Institute of Education and Child Studies, Leiden University, Leiden, Netherlands

OPEN ACCESS

Edited by:

Bert De Smedt,
KU Leuven, Belgium

Reviewed by:

Koen Luwel,
KU Leuven, Belgium
Katherine M. Robinson,
University of Regina, Canada

*Correspondence:

Marian Hickendorff
hickendorff@fsw.leidenuniv.nl

Specialty section:

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

Received: 14 March 2018

Accepted: 16 August 2018

Published: 11 September 2018

Citation:

Fagginger Auer MF, Hickendorff M
and van Putten CM (2018) Training
Can Increase Students' Choices
for Written Solution Strategies
and Performance in Solving
Multi-Digit Division Problems.
Front. Psychol. 9:1644.
doi: 10.3389/fpsyg.2018.01644

Making adaptive choices between solution strategies is a central element of contemporary mathematics education. However, previous studies signal that students make suboptimal choices between mental and written strategies to solve division problems. In particular, some students of a lower math ability level appear inclined to use mental strategies that lead to lower performance. The current study uses a pretest-training-posttest design to investigate the extent to which these students' choices for written strategies and performance may be increased. Sixth graders of below-average mathematics level ($n = 147$) participated in one of two training conditions: an explicit-scaffolding training designed to promote writing down calculations or a practice-only training where strategy use was not explicitly targeted. Written strategy choices and performance increased considerably from pretest to posttest for students in both training conditions, but not in different amounts. Exploratory results suggest that students' strategy choices may also be affected by their attitudes and beliefs and the sociocultural context regarding strategy use.

Keywords: mathematics, multi-digit arithmetic, division, solution strategies, adaptivity, training

INTRODUCTION

Tasks are executed using a variety of strategies during all phases of development (Siegler, 1987, 2007; Shrager and Siegler, 1998). For example, infants vary in their use of walking strategies (Snapp-Childs and Corbetta, 2009), first graders in their use of spelling strategies (Rittle-Johnson and Siegler, 1999), and older children in their use of transitive reasoning strategies (Sijtsma and Verweij, 1999). This large variance in strategies goes together with widely differing performance rates of the different strategies, thereby having profound effects on performance levels. As such, strategies have received ample research attention.

Children's and adults' strategy use has been investigated for many cognitive tasks, such as mental rotation (Janssen and Geiser, 2010), class inclusion (Siegler and Svetina, 2006), and analogical reasoning (Stevenson et al., 2011). A cognitive domain that has featured prominently in strategy research is arithmetic. Many studies have been conducted on elementary addition (e.g., Geary et al., 2004; Barrouillet and Lépine, 2005), subtraction (e.g., Barrouillet et al., 2008), multiplication (e.g., Van der Ven et al., 2012), and division (e.g., Mulligan and Mitchelmore, 1997;

LeFevre and Morris, 1999; Campbell and Xue, 2001; Robinson et al., 2006), which concern operations in the number domain up to 100 that are taught in the lower grades of primary school. Fewer studies have addressed strategy use in more complex multidigit arithmetical tasks in the higher grades, involving larger numbers or decimal numbers (e.g., Selter, 2001; Van Putten et al., 2005; Torbeyns et al., 2009; Schulz and Leuders, 2018). Multidigit division in particular is an understudied topic. Since many students experience difficulties in this domain, further study into the strategies students use and how these are affected by student and instructional factors is called for (Hickendorff et al., 2010; Robinson, 2017).

Adaptive Strategy Use

Strategy use in both elementary and multidigit arithmetic consists of different components (Lemaire and Siegler, 1995): individuals' strategy repertoire (which strategies are used); frequency (how often each strategy is used); efficiency (the accuracy and speed of each strategy); and adaptivity (whether the most suitable strategy for a given problem is used). These four aspects together shape arithmetical performance.

With mathematics education reforms that have taken place in various countries over the past decades (Kilpatrick et al., 2001), *adaptive expertise* has become increasingly important (Baroody, 2003; Hatano, 2003; Verschaffel et al., 2009; McMullen et al., 2016). Adaptive expertise includes flexibility (using various strategies) and adaptivity (selecting the optimal strategy). It contrasts with *routine expertise*, where children apply standard procedures in an inflexible and inadaptive way (Hatano, 2003). Choosing the most suitable strategy for a given problem (i.e., making an adaptive strategy choice) is therefore crucial in contemporary mathematics education.

There are several ways to define adaptivity of a strategy choice, dependent on what is considered the most suitable or "optimal" strategy (Verschaffel et al., 2009). One way is to define adaptivity solely based on task variables: the characteristics of a problem determine which strategy is optimal (e.g., the adaptive strategy choice for a problem like $1089 \div 11$ would be to use the compensation strategy: $1100 \div 11 = 100$). However, individuals differ in their mastery of different strategies, and the strategy that is most efficient for one person does not have to be the most efficient strategy for another person. A second, more comprehensive, definition of adaptivity therefore also takes individual differences into account: the optimal strategy is the one that is most efficient for a given problem for a particular person. A third definition even includes contextual variables in the definition, such as aspects of the test (e.g., time restrictions and characteristics of preceding problems) and affective aspects of the broader sociocultural context.

Strategy use is not an exclusively cognitive endeavor. Affective factors, like individuals' beliefs, attitudes, and emotions toward mathematics in general and (adaptive) strategy use in particular, to some extent influenced by the sociocultural context, have been argued to be very important in shaping individuals' strategy repertoire and choices (Ellis, 1997; Verschaffel et al., 2009). Ellis (1997) identified several affective, sociocultural factors that impact strategy use. Students have an implicit understanding of

which ways of problem solving are valued by their community: whether speed or accuracy is more important; whether mental strategies are valued over using external aids; whether using conventional procedures or original approaches is preferred; and whether asking for help in problem solving is desirable.

Given the importance of affective variables (attitudes and beliefs) as determinants of (adaptive) strategy use and the scarcity of research addressing this, further research is called for. We argue that it is theoretically interesting as well as practically highly relevant to investigate in what way the sociocultural context may be manipulated to favorably influence strategy choices. A domain for which this is particularly relevant is multidigit division, since studies reported that students tend to make sub-optimal choices between mental and written strategies for this type of problems (Hickendorff et al., 2009, 2010; Fagginger Auer et al., 2016), which will be elaborated on in the following.

Strategies for Solving Multi-Digit Division Problems

In mathematics education reform, standard, digit-based written algorithms to solve multi-digit arithmetic problems have a less prominent role than in more traditional mathematics education (Torbeyns and Verschaffel, 2016). In the Netherlands, the traditional algorithm for the operation of division was even abandoned in favor of a new standardized strategy: the whole-number-based approach (Janssen et al., 2005; Buijs, 2008). The major difference between the digit-based algorithm and the whole-number-based approach is whether or not the place value of the digits in the numbers is ignored or respected (Hickendorff et al., 2017; see **Table 1** for examples). That is, in the digit-based algorithm the place value of the digits is ignored (e.g., in **Table 1**, the "54" of 544 is dealt with as 54 and not as 540), whereas the whole-number-based approach respects the place value (e.g., in **Table 1**, 340 is subtracted from 544; Van den Heuvel-Panhuizen et al., 2009). In contemporary mathematics textbooks, the whole-number-based approach is instructed from fifth grade onward, and it is not before sixth grade that the digit-based is instructed (Hickendorff et al., 2017).

Dutch national assessments in 1997 and 2004 showed a decrease in sixth graders' use of the digit-based algorithm, but use of the whole-number-based approach did not increase accordingly. Instead, students made more use of strategies without any written work (Hickendorff et al., 2009). These mental

TABLE 1 | Examples of the digit-based algorithm, whole-number-based approach, and other written strategies applied to the division problem $544 \div 34$.

Digit-based algorithm	Whole-number-based approach	Non-algorithmic strategies
$54/544 \setminus 16$	$544:34 =$	$10 \times 34 = 340$
<u>34</u>	<u>340</u> - $10 \times$	$13 \times 34 = 442$
204	204	$16 \times 34 = 544$
<u>204</u>	<u>102</u> - $3 \times$	
0	102	
	<u>102</u> - <u>3</u> $\times$ +	
	0 16 $\times$	

strategies turned out to be very inaccurate compared to written strategies (digit-based or otherwise), suggesting that suboptimal strategy choices were made. This partly explained the large performance decline that was observed for multidigit division in the assessments (Hickendorff et al., 2009).

In follow-up studies, Fagginger Auer et al. (2016) and Hickendorff et al. (2010) showed that performance improved when writing down calculations was required in (lower mathematical ability) students who spontaneously solved division problems without any written work. This shows that a contextual factor - requiring the use of more efficient strategies - can affect performance favorably in the short term. A valuable next step would be an investigation of instructional contexts that increase students' *spontaneous choices* for efficient strategies, thereby foregrounding improvements in performance in a more sustainable way than by using test instructions to force students to write down their work.

Present Study

The present study is intended as a first step of such an investigation. It focuses on (1) the determinants of students' spontaneous choices between mental and written division strategies and (2) the effect of a training designed to increase students' choices for written rather than mental strategies, and thereby also their performance. Using a pretest-training-posttest design, an explicit-scaffolding training condition designed to promote writing down calculations was compared to a practice-only training condition where strategy use was not explicitly targeted. The explicit-scaffolding training involved a step-by-step problem-solving plan for multi-digit division problems, based on the principles of direct, explicit instruction that lower-ability students tend to profit from (Kroesbergen and Van Luit, 2003; Gersten et al., 2009). The practice-only training involved practicing problem solving only, without explicit scaffolding, but with feedback on the accuracy of the outcome as in the explicit-scaffolding condition.

The study focuses on sixth graders of below-average mathematics achievement level. We focused on sixth graders since in the Netherlands instruction in standardized written strategies begins in grade five. Therefore, sixth graders are likely to have experience with written strategies which would be a prerequisite to choose them. After grade six students enter secondary school, where other aspects of mathematics are central to instruction and practice. We focused on below-average achievers because these students tend to be more inclined to use mental strategies than their higher-achieving peers, whereas they have the lowest performance with mental strategies (Hickendorff et al., 2010; Fagginger Auer et al., 2016). In other words: with these students there is most need for, as well as most room for, improvement.

The study aimed to address three sets of research questions and accompanying hypotheses. Research question 1 was: to what extent are individual differences in strategy choice (mental vs. written) related to students' attitudes and beliefs toward mathematics in general and toward strategies in particular, and to aspects of the sociocultural context of the students' mathematics classroom (mathematics instruction, teacher attitudes and

beliefs)? This investigation is exploratory in nature, and therefore we did not formulate *a priori* hypotheses.

Research question 2 was: to what extent do the two training types affect students' strategy choice? Hypothesis 2a was that written strategy choices increase more from pretest to posttest in the explicit-scaffolding training than in the practice-only training. Hypothesis 2b was that the effects of the explicit-scaffolding training on the use of written strategies is larger for boys than girls, since boys tend to use more mental strategies in division than girls (Hickendorff et al., 2009, 2010; Fagginger Auer et al., 2013).

Research question 3 was: to what extent do the two training types affect students' performance? Hypothesis 3a was that performance increases from pretest to posttest in both training types since students in both conditions can practice solving division problems and receive outcome feedback. Hypothesis 3b was that the performance increase in the explicit-scaffolding training is larger than in the practice-only training, as a corollary of the expected increase of written strategies in the former group. Furthermore, within the explicit-scaffolding training, we expect to find different performance gains with regard to students' gender, mathematical ability level and working memory capacity (hypothesis 3c–3e). Hypothesis 3c was that the performance gains are larger in boys than in girls, as a corollary of the expectation that boys show a larger increase in written strategies use (cf. hypothesis 2b). Hypothesis 3d was that performance gains are larger for students with lower compared to higher mathematical ability level, because mental strategies are especially inaccurate for lower ability students (Hickendorff et al., 2010; Fagginger Auer et al., 2016). Finally, Hypothesis 3e was that training has a larger effect on performance when students' working memory capacity is lower, since mental strategies demand working-memory resources. Freeing up those resources by writing down calculations may therefore have a larger impact in students with lower working-memory capacity (in line with cognitive load theory; Paas et al., 2003). This is especially relevant in our sample, given that students with a lower mathematical ability tend to have a lower working memory capacity than higher ability students (Friso-van den Bos et al., 2013).

MATERIALS AND METHODS

Participants

In total, 19 different classes of 15 different schools agreed to participate. The schools were located in different medium-sized to large cities in the megalopolis in central-west Netherlands (the Randstad) and from one smaller city in east Netherlands.

There were 323 sixth graders in total, of whom 186 students had a percentile score below 50 on the most recent standardized national mathematics test (Janssen et al., 2010). Furthermore, students with a percentile score below 10 ($n = 39$) were excluded because atypical problems such as dyscalculia could occur in this group. Our effective sample of below-average achievers (percentile score between 10 and 50) thus contained 147 students (64 percent girls; mean age 11 year 9 month with $SD = 5$ month). These students were assigned to one of the

two training conditions using random assignment with gender, ability quartile and school as blocking variables: 74 received explicit-scaffolding training and 73 practice-only training.

The 19 teachers of the students (8 female) were on average 38 years old. Four different textbooks were used across the classes: Wereld in Getallen (9 classes), Pluspunt (5 classes), Alles Telt (4 classes), and Rekenrijk (1 class).

Materials

Pretest and Posttest

The pretest to assess students' division strategy choices and performance contained twelve multidigit division problems presented in **Table 2**. These problems were selected from the two most recent national assessments of mathematical ability at the end of primary school (Janssen et al., 2005; Scheltens et al., 2013), so that they resemble the type of problems students are used to solving (ecological validity). All problems were situated in realistic problem solving context (e.g., determining how many bundles of 40 tulips can be made from 2500 tulips), except for the problem $31.2 \div 1.2$. The test also contained twelve problems involving other mathematical operations (all from the most recent national assessment) as filler items. The posttest was identical to the pretest to allow for a direct comparison of results. Since the pretest and posttest were a month apart and students are used to solve arithmetic problems on a daily basis in their mathematics lessons during that period, it was very unlikely that students remembered any of the (rather complex) solutions.

Prior to the pretest and the posttest students received an instruction in which the experimenter explained that the students had to do a booklet with mathematics problems. The researcher explicitly stated that this was not a test but that (s)he was interested in learning more about how students go about solving such problems. Furthermore, students were instructed that if they wanted to write down calculations, they could do so in the booklet.

After students completed the mathematics problems in the booklets, the accuracy of the answer (correct or incorrect) and use of written work (yes or no) were scored for each problem. Solutions with written work were further classified into one of

three strategy categories: the digit-based algorithm, the whole-number-based approach, and other written strategies (see **Table 1** for examples).

Training Problems

The problems used in the three training sessions between the pretest and posttest were three sets of parallel versions of the twelve problems in **Table 2**.

Student and Teacher Questionnaires

The students filled out a questionnaire of seven questions (**Appendix A**) on their attitudes and beliefs toward mathematics in general and strategies in particular. The teachers filled out a questionnaire of fifteen questions (**Appendix B**) on their instructional practices regarding standardized division strategies, and attitudes/beliefs toward the importance of writing down calculations and various aspects of flexible and adaptive strategy use. The student and teacher questionnaires were devised specifically for this study.

Working Memory Tests

Students' working memory capacity was assessed using a computerized version of the digit span test from the WISC-III (Stevenson and De Bot, unpublished; Wechsler, 1991), and their spatial working memory using a computerized version of the Corsi block test (Corsi, 1972).

Training

In the training sessions, students worked on the set of training problems for that week. The experimenter evaluated each answer when it was written down and told the student whether it was correct or incorrect. When correct, the students proceeded to the next problem. When incorrect, the student tried again. Accuracy feedback was provided again, and regardless of whether the solution was correct this time, the student proceeded to the next problem. The session was terminated when 15 min had passed.

Two aspects differed between the two training types. First, students in the practice-only training were free in how they solved the problems (just as in the pretest), whereas the students in the explicit-scaffolding condition had to write down their calculations in a way that "would allow another child to see how they had solved the problem" (but apart from that, the choice for which type of written strategies was free). Second, when students in the practice-only condition failed to provide the correct answer in their first problem-solving attempt, they did not receive any feedback other than that the answer was incorrect before they could try to solve the problem in the second attempt. By contrast, when students in the explicit-scaffolding condition failed to provide the correct answer the first time, they were provided with explicit systematic scaffolding how to write down their calculations in a standardized way at the second attempt. A printed version of this step-by-step plan was always on the table so that students could use it whenever they wanted. When students were stuck in their problem solving, the experimenter used the plan and standardized verbal instructions to help the students with writing down calculations. No feedback was given

TABLE 2 | The division problems in pretest and posttest. Problems presented in *italics* are parallel versions of the problems that are not yet released for publication.

Number	Problem
1.	$1536 \div 16 = 96$
2.	$872 \div 4 = 218$
3.	$31.2 \div 1.2 = 26$
4.	$6496 \div 14 = 464$
5.	$544 \div 34 = 16$
6.	$11585 \div 14 = 827.5$
7.	$47.25 \div 7 = 6.75$
8.	$157.50 \div 7.50 = 21$
9.	$2500 \div 40 = 62$
10.	$1470 \div 12 = 122.50$
11.	$736 \div 32 = 23$
12.	$16300 \div 420 = 39$

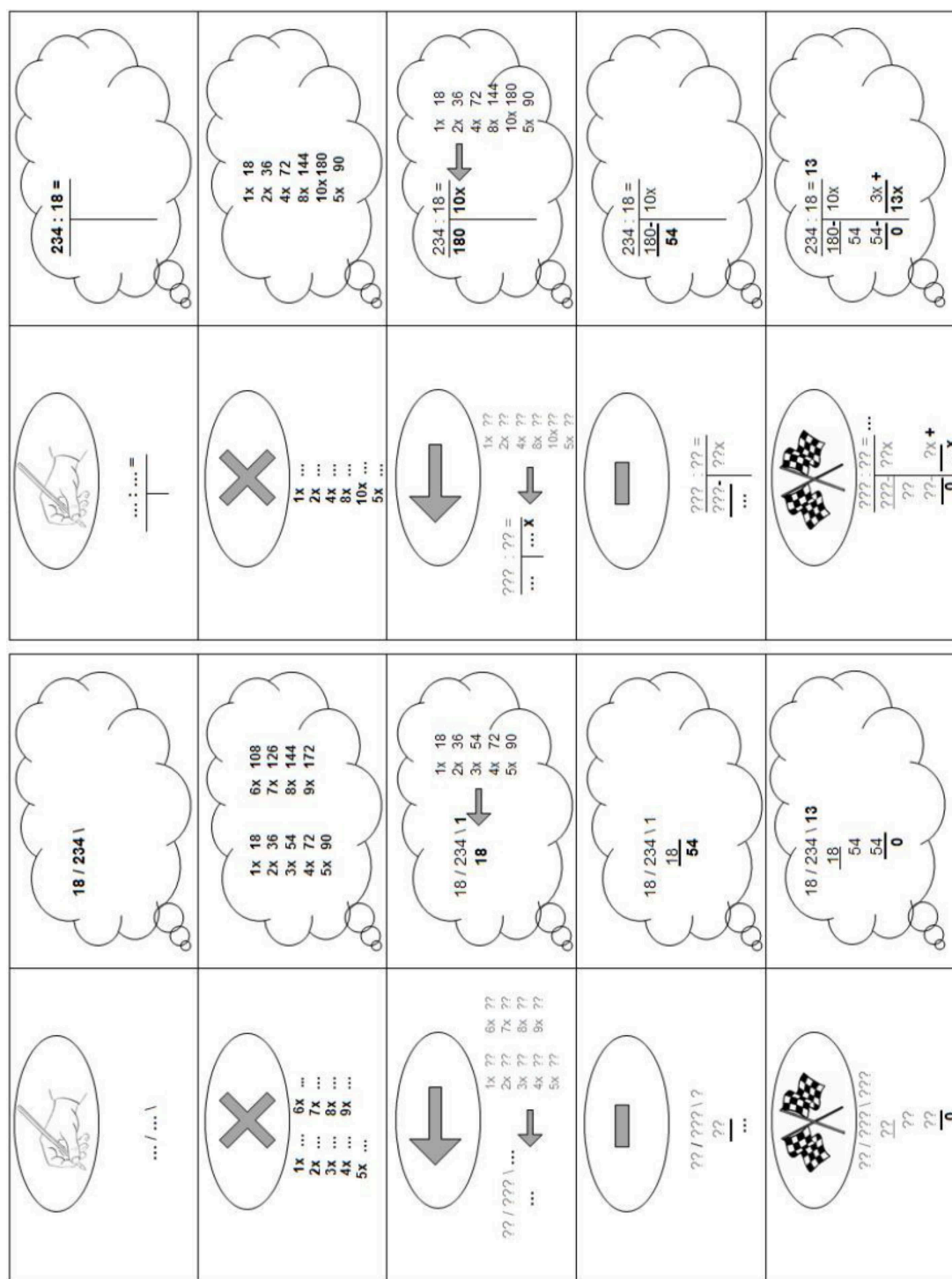


FIGURE 1 | The step-by-step plans in the explicit-scaffolding training in two versions: for students using the digit-based algorithm, and for students using the whole-number-based approach.

on the accuracy of what students wrote down (e.g., mistakes in the multiplication table), except for the final answer.

Since classes differed in which type of standardized strategy was instructed, there were two versions of the plan: one for

students taught the digit-based algorithm and one for students taught the whole-number-based approach (see **Figure 1**). In cases where students were taught both standardized strategies, the experimenter showed both step-by-step plans and the student

could select the strategy (s)he was used to applying. Both versions consist of five highly similar steps (with step 3 and 4 repeated as often as necessary): (1) writing down the problem; (2) writing down a multiplication table (optional step); (3) writing down a number (possibly from that table) to subtract; (4) writing down the subtraction of that number; and (5) finishing when zero is reached, which in the case of the whole-number-based approach requires a final addition of the repeated subtractions. Each step was represented by a symbol to make the step easy to identify and remember (the symbols in the ellipses on the left side of the scheme). Below this symbol, a general representation of the step was given, with question marks for problem-specific numbers already present at that step and dots for the numbers to be written down in that step. On the right-hand side of the plan, an example of the execution of each step for the particular problem $234 \div 18$ was given in a thinking cloud. On both sides, the elements to be written down in the current step were in bold font.

Procedure

The study was conducted over a period of 5 weeks in the fall. In week 1, students first completed the pretest in their classroom, in a maximum of 45 min, and also the two working memory tasks (on the computer) and the student questionnaire. In week 2–4, students participated in three individual training sessions of 15 min each (one per week) with an experimenter. The experiment was concluded in week 5, in which students completed the posttest. The teacher filled out the teacher questionnaire in week 1.

Statistical Analysis

Research Question 1

Correlations were used to explore relations between students' percentage of written strategy choices across the twelve pretest problems on the one hand and (a) student factors (attitudes and beliefs, based on student questionnaire) and (b) classroom factors (mathematics educational practices and sociocultural context, based on teacher questionnaire) on the other. These were point-biserial correlations for dichotomous questionnaire responses and Spearman's rank correlations for scales.

Research Questions 2 and 3

Explanatory IRT models (De Boeck and Wilson, 2004) were used to model the effect of the training types on pretest-posttest differences in strategy choice (question 2) and in performance (question 3), as well to investigate differential training effects by students' gender, mathematical ability level, and working memory. Measuring learning and change has inherent problems (Embretson and Reise, 2000; Stevenson et al., 2013). For instance, the interpretation of change scores depends on the score at pretest (e.g., a change from 1 to 3 may not mean the same as a change from 6 to 8), because sum scores in general and change scores in particular are not of interval measurement level. IRT models place persons and items on a common latent scale, resulting in a higher likelihood that the persons' ability estimates are of interval measurement level than simple sum scores (Embretson and Reise, 2000). To answer research

question 2, the dependent variable of the IRT models was strategy choice (written vs. not written) on each problem of the pretest and posttest, whereas it was accuracy of the answer (correct vs. incorrect) in the analyses to answer research question 3.

IRT models can be extended with an explanatory part by including explanatory variables, which can be item factors, person factors, and person-by-item factors. The current analyses included the following person factors: students' training condition, gender, mathematical ability score, and working memory. The person-by-item factor solution strategy choice (mental vs. written) was included in research question 3 only.

(Explanatory) IRT models can be estimated as multilevel logistic regression models, using general purpose software for generalized linear mixed models (GLMM) (De Boeck and Wilson, 2004). In the present study, the models were fitted using the lme4 package in R (De Boeck et al., 2011; Bates et al., 2014). All models were random person-random item Rasch models (RPRI; De Boeck, 2008), with a random intercept for students, and also a random intercept for items (as the problems were considered a draw from the larger domain of multidigit division). The explanatory variables were added in stepwise fashion (as in Stevenson et al., 2013, see also Paviyas et al., 2016), allowing evaluation of the added value of each step by comparing the models based on the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and likelihood ratio tests. The AIC and BIC balance model fit and parsimony (lower values are better). The likelihood ratio test (LRT) statistically tests the added value of including a specific explanatory variable by testing whether the more complex model with this specific explanatory variable included fits significantly better than the less complex model (without that variable).

For an indication of the size of significant effects, the probability of using a written strategy (question 2), or providing a correct answer (question 3), is computed for different levels of the explanatory variable of interest (with all other explanatory variables in the model set at the sample mean in the sample). For example, for the effect of testing occasion (pretest or posttest), the probability of a correct answer for an average student on an average problem on both the pretest and the posttest is computed. For scale variables (e.g., mathematics ability score) the effects of a difference of one standard deviation around the mean ($M - 0.5SD$ to $M + 0.5SD$) are given.

RESULTS

Research Question 1: Determinants of Written Strategy Choices

Students used written strategies in 59 percent of their pretest solutions, which varied across problems between 33 percent ($31.2 \div 1.2$) and 76 percent ($54.4 \div 3.4$), and across students between 0% ($n = 13$) and 100% ($n = 15$). In the following we report correlations between students' percentage of written strategy choices across the twelve

pretest problems on the one hand, and (a) student factors (attitudes and beliefs) and (b) classroom factors (mathematics educational practices and sociocultural context) on the other.

Student Factors

Appendix A shows the frequencies of the students' ($n = 147$) responses to the questionnaire items regarding their mathematical attitudes and beliefs, and the correlation of these responses with their percentage of written strategy choices at pretest. In the following, only significant correlations are discussed.

On average, the students had a slightly positive attitude toward mathematics ($M = 2.8$ on a 5-point scale), reported putting quite some effort into math ($M = 4.3$), were slightly positive about their mathematical ability ($M = 2.8$), and almost all students (97 percent) reported valuing accuracy over speed. These factors were not significantly related to written strategy choices. On the questions concerning strategy use, a majority of students (77 percent) found it more important to be able to solve mathematical problems with rather than without paper, and this was positively related to using written strategies ($r = 0.23$). Students reported that they sometimes answer without writing down a calculation ($M = 2.6$) and this self-reported frequency of non-written strategies was negatively related to using written strategies at pretest ($r = -0.19$). When asked to select their reasons to not write down calculations (multiple answers possible), the most frequently reported reason (selected by 49 percent of students) was because they “did not feel it was necessary,” followed by because “it was faster” (41 percent). Other reasons were because of “not feeling like it” (22 percent), because they “guessed the solution rather than computing it” (14 percent), because “mental strategies are more accurate” (14 percent), and because “it is smarter to be able to solve a problem mentally” (11 percent). Virtually no students (1 percent) reported they used a mental strategy because “it was cooler.”

Classroom Factors

Appendix B shows what the 19 teachers reported on the teacher questionnaire. With the exception of one item, the teachers' responses were unrelated to their students' use of written strategies. Most teachers taught their students only the whole-number-based approach exclusively ($n = 11$) or in combination with the digit-based algorithm ($n = 5$); three teachers taught their students the digit-based algorithm exclusively. On average, teachers did not prefer one standardized strategy over the other ($M = 3.0$), but did prefer the use of standardized over non-standardized approaches ($M = 2.2$). Only this item correlated with students' use of written strategies: the more teachers preferred non-standardized strategies, the lower the percentage of their students' written strategies ($r = -0.46$). On average, teachers found performing calculations well on paper and mentally equally important for their students ($M = 3.0$). They reported instructing their students in writing down calculations frequently (on average almost daily, $M = 4.2$). Concerning multidigit division problems specifically, teachers on average

found writing down calculations somewhat more important for their students than trying to do it mentally ($M = 2.4$) and valued accuracy somewhat over speed ($M = 2.5$). Making a good estimation of the solution was valued more than being able to determine the exact solution ($M = 3.5$), as was knowing more solution procedures rather than just one ($M = 3.4$). Teachers considered using a standardized approach versus choosing a custom solution strategy on average equally important ($M = 3.0$), and valued convenient shortcut strategies somewhat more than using an approach that can always be applied ($M = 3.3$).

Research Questions 2 and 3

Descriptive Statistics

Table 3 presents descriptive statistics about the content of the training. As instructed, students in the explicit-scaffolding condition virtually always wrote down a calculation (98–99 percent). Though not instructed to do so, students in the practice-only condition also had a high and increasing tendency to use written strategies (81–93 percent). The feedback in the explicit-scaffolding condition (on average 3.3 times per session) included writing down a multiplication table (0.8 times), selecting a number from that table (1.1 times), writing down of the problem (0.5 times), subtracting the selected number (0.5 times), and finishing the procedure (0.5 times).

Research Question 2

The effects of the training on written strategy choices were evaluated using a series of explanatory IRT models on the pretest and posttest data, with successively more explanatory variables (see **Table 4**). First a baseline model for the probability of a written strategy choice was fitted with only random intercepts for students and problems and no covariates (model M_0). In model M_1 , main effects were added for the student characteristics gender, mathematical ability and working memory capacity, which improved fit according to all criteria. Fit was further improved by adding a main effect for testing occasion (pretest vs. posttest; model M_2). However, the change in written strategy choices from pretest to posttest did not significantly differ for the two training groups (model M_3). Adding interactions between condition, testing occasion and student characteristics also did not improve the model (models are not included in **Table 4**).

Interpretation of the best fitting model, M_2 , shows that girls used more written strategies ($P = 0.94$) than boys ($P = 0.74$), $z = -6.0$, $p < 0.001$, and that mathematical ability score was positively associated with using written strategies ($P = 0.80$ vs. $P = 0.92$ for one standard deviation difference), $z = 4.3$, $p < 0.001$. Working memory (sum score of the verbal and spatial working memory scores) had no significant effect, $z = -0.6$, $p = 0.55$. Students used more written strategies at the posttest ($P = 0.94$) than at the pretest ($P = 0.76$), $z = 13.5$, $p < 0.001$.

To investigate whether the two trainings differ in the type of written strategies they elicited, **Table 5** presents a more detailed categorization of strategies than just written or non-written. It shows that the frequency of using the

TABLE 3 | Descriptive statistics of training sessions (averages across students).

Training	Number of problems per session	Number of second attempts per session	Feedback Frequency per session	% Written strategies		
				Session 1	Session 2	Session 3
Explicit scaffolding	5.1	1.6	3.3	98	99	99
Practice-only	6.1	1.8	–	81	87	93

TABLE 4 | Explanatory IRT models for effects on written strategy choices (all comparisons are to M_{n-1}).

Model	Added fixed effects	LL	# Parameters	AIC	BIC	Likelihood Ratio Test
M_0		–1337.6	3	2681.1	2699.4	
M_1	Gender, math ability, and working memory	–1315.7	6	2643.3	2679.8	$\chi^2(3) = 43.8, p < 0.001$
M_2	Testing occasion	–1216.5	7	2447.0	2489.5	$\chi^2(1) = 198.3, p < 0.001$
M_3	Condition $\times$ occasion	–1215.6	9	2449.2	2503.9	$\chi^2(2) = 1.7, p = 0.42$

digit-based algorithm and whole-number-based approach, other written strategies, non-written strategies and other strategies is almost identical (differences of no more than five percentage points) in the two training groups - both at pretest and at posttest. In both groups, similar increases in the use of both types of standardized strategies and decreases in the use of other written and non-written strategies occurred.

Research Question 3

Model fit statistics for performance (accuracy) are presented in **Table 6**. As for written strategy choices, first a baseline model for the probability of a correct response was fitted (M_0), and again, this model was improved by adding student gender, ability and working memory (M_1) and by adding testing occasion (M_2), but not by adding condition effects (M_3). The best fitting model, M_2 , shows that girls ($P = 0.43$) performed better than boys ($P = 0.28$), $z = -3.8, p < 0.001$, and that general mathematics ability score was positively associated with performance ($P = 0.28$ vs. $P = 0.43$ for one SD difference), $z = 4.5, p < 0.001$. Working memory had no significant effect, $z = 0.04, p = 0.97$. Students performed better at the posttest ($P = 0.48$) than at the pretest ($P = 0.24$), $z = 11.9, p < 0.001$.

Next, the difference in accuracy between written and non-written strategies was investigated by fitting a model for accuracy with main effects for all previous predictors (student characteristics, testing occasion, and condition) and strategy choice (written or not), and all first-order interactions between strategy choice and the other predictors. This showed that written strategies were much more accurate ($P = 0.40$) than non-written strategies ($P = 0.19$), $z = 4.1, p < 0.001$, and that this did not depend significantly on testing occasion, $z = 1.1, p = 0.27$, gender, $z = 0.0, p = 0.99$, ability, $z = 1.0, p = 0.32$, working memory, $z = 0.3, p = 0.75$, or condition, $z = -1.0, p = 0.33$. Finally, we investigated the extent to which individual students' gains in written strategy choices from pretest to posttest were related to their gains in accuracy from pretest to posttest. Spearman's rank correlation between difference in written strategy use and difference in accuracy was significant

TABLE 5 | Strategy use proportions on the pretest and posttest in the different training conditions.

	Pretest		Posttest	
	Explicit-scaffolding	Practice-only	Explicit-scaffolding	Practice-only
Digit-based algorithm	0.09	0.09	0.13	0.13
Whole-number approach	0.37	0.40	0.61	0.62
Other written	0.19	0.19	0.13	0.08
No written work	0.35	0.30	0.13	0.17
Remainder	0.01	0.02	0.00	0.00

and positive: $r(142) = 0.23, p = 0.006$. These results show that not only written strategies are more accurate than mental ones, but also that increasing the use of written strategies leads to increased performance.

DISCUSSION

The current study's aim was to investigate determinants of below-average sixth graders' choices between mental and written strategies for solving multi-digit division problems, and the effect of a training to increase students' choices for written rather than mental strategies. First, exploratory analyses showed that individual differences in strategy choice (mental vs. written) were related to some aspects of students' attitudes and beliefs toward strategy use, but not to their attitudes and beliefs toward mathematics in general. Specifically, students who reported that it is more important to solve problems with rather than without paper, and students who reported not so often using non-written strategies were more inclined to use written strategies at the pretest items. Students' individual differences in strategy choice were related to only one aspect of the sociocultural context (as measured with a teacher questionnaire): the more teachers valued standardized over non-standardized strategies, the more their students used written strategies. An important remark is that since there were only 19 teachers in our sample, low statistical power may have prevented finding other significant

TABLE 6 | Explanatory IRT models for effects on accuracy (all comparisons are to M_{n-1}).

Model	Added fixed effects	LL	# Parameters	AIC	BIC	Likelihood Ratio Test
M_0		-1801.0	3	3607.9	3626.1	
M_1	Gender, math ability, and working memory	-1785.3	6	3582.5	3619.0	$\chi^2(3) = 31.4, p < 0.001$
M_2	Testing occasion	-1711.1	7	3436.3	3478.8	$\chi^2(1) = 148.3, p < 0.001$
M_3	Condition $\times$ occasion	-1710.8	9	3439.6	3494.2	$\chi^2(2) = 0.7, p = 0.70$

associations. Furthermore, the students were instructed by their current teacher for only 2–4 months, which could be another explanation that there were hardly any relations found between teachers' instructional practices and students' strategy use. Overall, teachers reported frequent instruction in writing down calculations, preferred use of a standardized over a non-standardized strategy, and valued written strategies somewhat over mental strategies and accuracy somewhat over speed. These results suggest a sociocultural context in which there is room for written strategies, but where it is not the highest priority.

In the second part of the study, the effects of a training designed to promote students' choices for written rather than mental strategies (and thereby, their performance) were compared to the effects of a practice-only training. In both training conditions the use of written strategies and accuracy increased from pretest to posttest, written strategies were more accurate than mental ones, and individual students' increase in the use of written strategies was related to their performance gains. However, the hypothesized differential training effects were not observed. Students' written strategy choices increased to the same extent in both training conditions (in contrast with hypothesis 2a) and there were no differential training effects for boys and girls (in contrast to hypothesis 2b). Regarding performance, performance (accuracy) increased in both groups from pretest to posttest (in line with hypothesis 3a), but not more so in the explicit-scaffolding training condition (in contrast to hypothesis 3b). Furthermore, there were no differential performance gains by gender, mathematical ability level, or working memory (in contrast to hypotheses 3c–3e).

All in all, written strategy choices and performance were considerably higher after training than before training, irrespective of the type of training. Both training types were thus effective in increasing the use of written strategies and thereby performance. However, the elements of explicit scaffolding written strategy use did not add to the effect of only practicing solving the problems with outcome feedback. While writing down calculations was not required during practice-only training, it did occur frequently and increasingly across the training sessions. In the first session calculations were written down in 81 percent of the problems - considerably more than the 70 percent during the pretest. This increased up to 93 percent in the third training session, whereas it decreased to 87 percent in the posttest again. As such, students practiced written calculations almost as much in the practice-only training as in the explicit-scaffolding training, reducing the contrast between the two conditions. The

common elements of both trainings – practicing written strategies with outcome feedback – therefore seem to account for the observed changes in strategy choices and accuracy.

In the practice-only condition, the relatively high frequency of written strategy choices in the training sessions compared to the pretest and posttest may possibly be explained by differences in the setting: in a classroom (at pretest and posttest) versus one-on-one with an experimenter (training sessions). Previous research showed a similar difference between a classroom administration setting and individual testing (Van Putten and Hickendorff, 2009). A possible explanation is that students use written strategies because they think the experimenter may expect or prefer that (i.e., demand characteristics; Orne, 1962), in line with the students' teachers' light preference of written over mental strategies.

The increase in the use of written strategies over the three training sessions in the practice-only training may possibly be explained by the direct accuracy feedback after each solution (Ellis et al., 1993), and the requirement to do a problem again when the first solution was incorrect. Direct accuracy feedback allows for an immediate evaluation of the success of the strategy that was applied, and this evaluation should often be in favor of written rather than mental strategies given the considerably higher accuracy of the former. Combined with the extra effort associated with an incorrect solution (redoing the problem), this is likely to be an important incentive for written strategy choices.

The element that was unique to the explicit-scaffolding training was the requirement to use a written strategy, scaffolded by a step-by-step plan for writing down calculations. The finding that this element apparently did not have an additional effect contrasts with the results of a meta-analysis on mathematics interventions for low-ability students that identified such plans as an important component of effective interventions (Gersten et al., 2009). In the current study students turned out to require little feedback based on the plan, and the feedback that was given most often concerned an optional element: the multiplication table. Furthermore, students in the practice-only training turned out to practice solving on average one problem more compared to students in the explicit-scaffolding training, which may have masked potential positive effects of the scaffolding elements (similar to Van de Pol et al., 2015).

In addition to the finding that there were no differences in the effects of the two training types, also no differential training effects by gender, mathematical ability and working memory were found. This may be explained by the same reasoning: in practice

the difference between the two training types may have been much smaller than intended.

Limitations

There are several limitations that deserve attention. First, there was no genuine control group of students who did not receive training. Therefore it is not possible to ascribe with certainty the gains in written strategy use and performance to the training. We did, however, collect pretest and posttest data from the 137 students with above-average mathematics achievement level who were in the participating classes, but did not participate in any of the trainings. The pretest-to-posttest increase in both the use of written strategies and in performance was significantly higher in the (below-average achieving) students who received training than in the (above-average) students who did not receive training. This differential learning effect supports confidence in the interpretation that it was the training that was effective in increasing written strategy use and performance, although the difference in achievement level between the two groups (below-average vs. above-average) possibly confounds this effect.

A second limitation is that there was no retention test. It was therefore not possible to analyze the stability of the trainings' effects. Future studies should include a follow-up test later in the school year to address this specifically.

A third limitation concerns the measurement of the teacher's instructional practices. The use of a questionnaire may not present a complete picture of the actual instructional activities taking place in the mathematics classroom (Porter, 2002), and future studies should include classroom observations to measure the instruction in a more direct way. Moreover, the amount of time the students were instructed by their teacher was relatively short (2–4 months) possibly weakening the effect the teacher's instructional practices may have had. Future studies could be conducted in the second half of the school year so that the students have received instruction from their teacher for a longer period of time.

Future Directions

The results of the present study provide several suggestions for future research on strategy training programs. The results suggest that direct accuracy feedback (possibly with some cost attached to incorrect solutions) may be conducive to beneficial changes in strategy choices. They also show that considerable changes in strategy choices and improvements in performance may be achieved with as few as three training sessions of 15 min (in line with the finding of Kroesbergen and Van Luit, 2003, who found that longer mathematics interventions are not necessarily more effective). As said, a follow-up test after a longer period of time (e.g., several months) should be used to establish whether the changes are lasting.

The results also provide suggestions for other possible ways to influence students' choices between mental and written strategies. Since strategy choices appear to be related to students' valuing of written strategies and to teachers' valuing of standardized over non-standardized strategies, a sociocultural context that highlights these aspects may affect strategy students' strategy choice (Ellis, 1997). This might be achieved by having teachers

express more appreciation of the use of external aids in problem solving and of standardizing written solution steps.

CONCLUSION

The present study showed that three training sessions in which students practice solving division problems with written strategies and receive feedback on the accuracy of the outcome, whether or not explicitly scaffolded with a step-by-step direct-instruction plan, increased below-average sixth graders' use of written strategies and performance in solving multi-digit division problems. Given the fact that students seem to make sub-optimal choices for non-written strategies in this domain, this is an important starting point for efforts to increase the use of written strategies. Further research is necessary to identify the optimal set-up of a training targeting students' written strategy use.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of ethical guidelines of the Ethics Committee of the Institute of Psychology, Leiden University. The protocol was approved by the Leiden University Psychology Research Ethics Committee (CEP number 6520034071). All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

MFA, MH, and CvP contributed to the design of this study. MFA organized the data collection and database, performed the statistical analyses, and wrote the first draft of the manuscript. MH wrote a major revision of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was supported by the Netherlands Organisation for Scientific Research (NWO) in the project "Mathematics instruction in the classroom and students' strategy use and achievement in primary education" with project number 411-10-706.

ACKNOWLEDGMENTS

We are indebted to CITO (Dutch National Institute of Educational Measurement) for giving the opportunity to use mathematics items from their national assessment tests. Furthermore, we thank Chris Hoeboer and Leonore Braggaaar for their assistance in conducting the study. Finally, we thank Claire Stevenson for the automated version of the Corsi block task.

REFERENCES

- Baroody, A. J. (2003). "The development of adaptive expertise and flexibility: the integration of conceptual and procedural knowledge," in *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise*, eds A. J. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates), 1–33.
- Barrouillet, P., and Lépine, R. (2005). Working memory and children's use of retrieval to solve addition problems. *J. Exp. Child Psychol.* 91, 183–204. doi: 10.1016/j.jecp.2005.03.002
- Barrouillet, P., Mignon, M., and Thevenot, C. (2008). Strategies in subtraction problem solving in children. *J. Exp. Child Psychol.* 99, 233–251. doi: 10.1016/j.jecp.2007.12.001
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear Mixed-Effects Models Using Eigen and S4*. Available at: <http://CRAN.R-project.org/package=lme4>
- Buijs, C. (2008). *Leren Vermenigvuldigen met Meercijferige Getallen [Learning to Multiply with Multidigit Numbers]*. Utrecht: Freudenthal Institute for Science and Mathematics Education.
- Campbell, J. I., and Xue, Q. (2001). Cognitive arithmetic across cultures. *J. Exp. Psychol. Gen.* 130, 299–315. doi: 10.1037/0096-3445.130.2.299
- Corsi, P. (1972). Human memory and the medial temporal region of the brain. *Diss. Abstr. Int.* 34:891.
- De Boeck, P. (2008). Random item models. *Psychometrika* 73, 533–559. doi: 10.1007/s11336-008-9092-x
- De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., et al. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *J. Stat. Softw.* 39, 1–28. doi: 10.18637/jss.v039.i12
- De Boeck, P., and Wilson, M. (eds). (2004). *Explanatory item Response Models: A Generalized Linear and Nonlinear Approach*. New York, NY: Springer. doi: 10.1007/978-1-4757-3990-9
- Ellis, S. (1997). Strategy choice in sociocultural context. *Dev. Rev.* 17, 490–524. doi: 10.1006/drev.1997.0444
- Ellis, S., Klahr, D., and Siegler, R. (1993). "Effects of feedback and collaboration on changes in children's use of mathematical rules," in *Paper Presented at the Biennial Meeting of the Society for Research in Child Development*, New Orleans, LA.
- Embretson, S. E., and Reise, S. P. (2000). *Item Response theory for Psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fagginger Auer, M. F., Hickendorff, M., and Van Putten, C. M. (2013). "Strategiegebruik bij het oplossen van vermenigvuldig- en deelopgaven [Strategy use in solving multiplication and division problems]," in *Balans van Het Reken-Wiskundeonderwijs aan Het einde van de Basisschool*, eds F. Scheltens, B. Hemker, and J. Vermeulen (Arnhem: CITO), 158–167.
- Fagginger Auer, M. F., Hickendorff, M., and Van Putten, C. M. (2016). Solution strategies and adaptivity in multidigit division in a choice/no-choice experiment: student and instructional factors. *Learn. Instruct.* 41, 52–59. doi: 10.1016/j.learninstruc.2015.09.008
- Friso-van den Bos, I., Van der Ven, S. H. G., Kroesbergen, E. H., and Van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: a meta-analysis. *Educ. Res. Rev.* 10, 29–44. doi: 10.1016/j.edurev.2013.05.003
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., and DeSoto, M. C. (2004). Strategy choices in simple and complex addition: contributions of working memory and counting knowledge for children with mathematical disability. *J. Exp. Child Psychol.* 88, 121–151. doi: 10.1016/j.jecp.2004.03.002
- Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., and Flojo, J. (2009). Mathematics instruction for students with learning disabilities: a meta-analysis of instructional components. *Rev. Educ. Res.* 79, 1202–1242. doi: 10.3102/0034654309334431
- Hatano, G. (2003). "Foreword," in *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise*, eds A. J. Baroody and A. Dowker (Mahwah, NJ: Lawrence Erlbaum Associates), xi–xiii.
- Hickendorff, M., Heiser, W. J., van Putten, C. M., and Verhelst, N. D. (2009). Solution strategies and achievement in Dutch complex arithmetic: latent variable modeling of change. *Psychometrika* 74, 331–350. doi: 10.1007/s11336-008-9074-z
- Hickendorff, M., Torbeyns, J., and Verschaffel, L. (2017). Grade-related differences in strategy use in multidigit division in two instructional settings. *Br. J. Dev. Psychol.* 36, 169–187. doi: 10.1111/bjdp.12223
- Hickendorff, M., Van Putten, C. M., Verhelst, N. D., and Heiser, W. J. (2010). Individual differences in strategy use on division problems: mental versus written computation. *J. Educ. Psychol.* 102, 439–452. doi: 10.1037/a0018177
- Janssen, A. B., and Geiser, C. (2010). On the relationship between solution strategies in two mental rotation tasks. *Learn. Individ. Dif.* 20, 473–478. doi: 10.1016/j.lindif.2010.03.002
- Janssen, J., Van der Schoot, F., and Hemker, B. (2005). *Balans van het Reken-Wiskundeonderwijs aan Het einde van de Basisschool 4 [Fourth Assessment of Mathematics Education at the end of Primary School]*. Arnhem: CITO.
- Janssen, J., Verhelst, N., Engelen, R., and Scheltens, F. (2010). *Wetenschappelijke Verantwoording van de Toetsen LOVS Rekenen-Wiskunde voor Groep 3 tot en Met Groep 8 [Technical Report for the Student Monitoring System Mathematics Tests for Grade 1 to 6]*. Arnhem: CITO.
- Kilpatrick, J., Swafford, J., and Findell, B. (2001). *Adding it up. Helping Children Learn Mathematics*. Washington, DC: National Academy Press.
- Kroesbergen, E. H., and Van Luit, J. E. H. (2003). Mathematics interventions for children with special educational needs: a meta-analysis. *Remedial Spec. Educ.* 24, 97–114. doi: 10.4103/2277-9531.162372
- LeFevre, J. A., and Morris, J. (1999). More on the relation between division and multiplication in simple arithmetic: evidence for mediation of division solutions via multiplication. *Mem. Cogn.* 27, 803–812. doi: 10.3758/BF03198533
- Lemaire, P., and Siegler, R. S. (1995). Four aspects of strategic change: contributions to children's learning of multiplication. *J. Exp. Psychol. Gen.* 124, 83–97. doi: 10.1037/0096-3445.124.1.83
- McMullen, J., Brezovszky, B., Rodríguez-Aflecht, G., Pongsakdi, N., Hannula-Sormunen, M. M., and Lehtinen, E. (2016). Adaptive number knowledge: exploring the foundations of adaptivity with whole-number arithmetic. *Learn. Individ. Dif.* 47, 172–181. doi: 10.1016/j.lindif.2016.02.007
- Mulligan, J. T., and Mitchelmore, M. C. (1997). Young children's intuitive models of multiplication and division. *J. Res. Math. Educ.* 28, 309–330. doi: 10.2307/749783
- Orne, M. T. (1962). On the social psychology of the psychological experiment: with particular reference to demand characteristics and their implications. *Am. Psychol.* 17, 776–783. doi: 10.1037/h0043424
- Paas, F., Renkl, A., and Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educ. Psychol.* 38, 1–4. doi: 10.1207/S15326985EP3801_1
- Pavias, M., van den Broek, P., Hickendorff, M., Beker, K., and van Leijenhof, L. (2016). Effects of social-cognitive processing demands and structural importance on narrative recall: differences between children, adolescents, and adults. *Discourse Process.* 53, 488–512. doi: 10.1080/0163853X.2016.1171070
- Porter, A. C. (2002). Measuring the content of instruction: uses in research and practice. *Educ. Res.* 31, 3–14. doi: 10.3102/0013189X031007003
- Rittle-Johnson, B., and Siegler, R. S. (1999). Learning to spell: variability, choice, and change in children's strategy use. *Child Dev.* 70, 332–348. doi: 10.1111/1467-8624.00025
- Robinson, K. M. (2017). "The understanding of additive and multiplicative arithmetic concepts," in *Acquisition of Complex Arithmetic Skills and Higher-Order Mathematics Concepts*, eds D. C. Geary, D. B. Berch, R. J. Ochsendorff, and K. M. Koepke (New York, NY: Elsevier Inc.), 21–46.
- Robinson, K. M., Arbuthnott, K. D., Rose, D., McCarron, M. C., Globa, C. A., and Phonexay, S. D. (2006). Stability and change in children's division strategies. *J. Exp. Child Psychol.* 93, 224–238. doi: 10.1016/j.jecp.2005.09.002
- Scheltens, F., Hemker, B., and Vermeulen, J. (2013). *Balans van Het Reken-Wiskundeonderwijs aan Het einde van de Basisschool 5 [Fifth Assessment of Mathematics Education at the end of Primary School]*. Arnhem: CITO.
- Schulz, A., and Leuders, T. (2018). Learning trajectories towards strategy proficiency in multi-digit division - a latent transition analysis of strategy and error profiles. *Learn. Individ. Dif.* 66, 54–69. doi: 10.1016/j.lindif.2018.04.014
- Selter, C. (2001). Addition and subtraction of three-digit numbers: German elementary children's success, methods and strategies. *Educ. Stud. Math.* 47, 145–173. doi: 10.1023/A:1014521221809
- Shrager, J., and Siegler, R. S. (1998). SCADS: a model of children's strategy choices and strategy discoveries. *Psychol. Sci.* 9, 405–410. doi: 10.1111/1467-9280.00076

- Siegler, R. S. (1987). The perils of averaging data over strategies: an example from children's addition. *J. Exp. Psychol. Gen.* 116, 250–264. doi: 10.1037/0096-3445.116.3.250
- Siegler, R. S. (2007). Cognitive variability. *Dev. Sci.* 10, 104–109. doi: 10.1111/j.1467-7687.2007.00571.x
- Siegler, R. S., and Svetina, M. (2006). What leads children to adopt new strategies? A microgenetic/cross-sectional study of class inclusion. *Child Dev.* 77, 997–1015. doi: 10.1111/j.1467-8624.2006.00915.x
- Sijtsma, K., and Verweij, A. C. (1999). Knowledge of solution strategies and IRT modeling of items for transitive reasoning. *Appl. Psychol. Measu.* 23, 55–68. doi: 10.1177/01466219922031194
- Snapp-Childs, W., and Corbetta, D. (2009). Evidence of early strategies in learning to walk. *Infancy* 14, 101–116. doi: 10.1080/15250000802569835
- Stevenson, C. E., Hickendorff, M., Resing, W., Heiser, W., and de Boeck, P. (2013). Explanatory item response modeling of children's change on a dynamic test of analogical reasoning. *Intelligence* 41, 157–168. doi: 10.1016/j.intell.2013.01.003
- Stevenson, C. E., Touw, K. W. J., and Resing, W. C. M. (2011). Computer or paper analogy puzzles: does assessment mode influence young children's strategy progression? *Educ. Child Psychol.* 28, 67–84.
- Torbeyns, J., Ghesquière, P., and Verschaffel, L. (2009). Efficiency and flexibility of indirect addition in the domain of multi-digit subtraction. *Learn. Instruct.* 19, 1–12. doi: 10.1016/j.learninstruc.2007.12.002
- Torbeyns, J., and Verschaffel, L. (2016). Mental computation or standard algorithm? Children's strategy choices on multi-digit subtractions. *Eur. J. Psychol. Educ.* 31, 99–116. doi: 10.1007/s10212-015-0255-8
- Van de Pol, J., Volman, M., Oort, F., and Beishuizen, J. (2015). The effects of scaffolding in the classroom: support contingency and student independent working time in relation to student achievement, task effort and appreciation of support. *Instruct. Sci.* 43, 615–641. doi: 10.1007/s11251-015-9351-z
- Van den Heuvel-Panhuizen, M., Robitzsch, A., Treffers, A., and Köller, O. (2009). Large-scale assessments of change in student achievement: Dutch primary school students' results on written division in 1997 and 2004 as an example. *Psychometrika* 74, 351–365. doi: 10.1007/s11336-009-9110-7
- Van der Ven, S. H. G., Boom, J., Kroesbergen, E. H., and Leseman, P. P. M. (2012). Microgenetic patterns of children's multiplication learning: confirming the overlapping waves model by latent growth modeling. *J. Exp. Child Psychol.* 113, 1–19. doi: 10.1016/j.jecp.2012.02.001
- Van Putten, C. M., and Hickendorff, M. (2009). Peilstokken voor plasterk: evaluatie van de rekenvaardigheid in groep 8 [Evaluating mathematics achievement in grade six]. *Tijdschr. Orthopedagogiek* 48, 183–194.
- Van Putten, C. M., Van den Brom-Snijders, P. A., and Beishuizen, M. (2005). Progressive mathematization of long division strategies in Dutch primary schools. *J. Res. Math. Educ.* 36, 44–73.
- Verschaffel, L., Luwel, K., Torbeyns, J., and Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *Eur. J. Psychol. Educ.* 24, 335–359. doi: 10.1007/BF03174765
- Wechsler, D. (ed.). (1991). *The Wechsler Intelligence Scale for Children*, 3rd Edn. San Antonio, TX: The Psychological Corporation.

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

The reviewer KL and handling editor declared their shared affiliation at the time of review.

Copyright © 2018 Fagginger Auer, Hickendorff and van Putten. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX A

Student Questionnaire

The proportion of students choosing each alternative is given in between brackets. For 5-point scales the mean is also presented. The correlations are between the students' question response and the students' frequency of written strategy choices on the pretest; statistically significant correlations are in bold.

- How much do you like math? *not at all (0.10)/not so much (0.21)/it's okay (0.45)/quite a bit (0.23)/a lot (0.01)*; $M = 2.85$; $r(144) = 0.06$, $p = 0.473$.
- How much effort do you put into math? *none (0.00)/not so much (0.01)/a bit (0.08)/quite a lot (0.53)/a lot (0.38)*; $M = 4.29$; $r(145) = 0.10$, $p = 0.236$.
- How good do you think you are at math? *not good at all (0.07)/not so good (0.25)/okay (0.46)/quite good (0.22)/very good (0.00)*; $M = 2.84$; $r(145) = 0.07$, $p = 0.385$.
- What is more important to you when you solve a mathematics problem? *solving the problem quickly (0.02)/finding the correct solution (0.98)*; $r(143) = 0.06$, $p = 0.487$.
- What is more important to you when you solve a mathematics problem? *being able to do it mentally (0.22)/being able to do it using paper (0.78)*; **$r(143) = 0.23$, $p = 0.005$** .
- How often do you solve problems without writing down a calculation? *almost never (0.16)/not often (0.29)/sometimes (0.40)/often (0.12)/very often (0.03)*; $M = 2.56$; **$r(145) = -0.19$, $p = 0.019$** .
- When you do not write down a calculation, why is that? (*tick boxes that apply*).
 - because it is faster (0.41)
 - because then you get a correct solution more often (0.14)
 - because doing mental calculation shows you are smart (0.11)
 - because it is cooler to do mental calculation (0.01)
 - because you do not feel like writing anything down (0.22)
 - because you guessed the solution (0.14)
 - because it is not necessary to write down a calculation (0.49)

APPENDIX B

Teacher Questionnaire

The proportion teachers choosing each alternative is presented between brackets. For 5-point scales the mean is also presented. The correlations are between the question response and the frequency of the teachers' students average percentage of written strategy choices on the pretest; statistically significant correlations are in bold.

- Do you teach your students the whole-number-based algorithm, digit-based algorithm or non-algorithmic approaches for solving multidigit problems (such as $544 \div 34$ or $12.6 \div 1.4$)? When multiple approaches apply, tick multiple boxes. *whole-number-based algorithm (0.58)/both whole-number-based and digit-based algorithm (0.26)/digit-based algorithm (0.16)*; $r(17) = -0.18$, $p = 0.462$.
- To what extent do you as a teacher prefer a division algorithm? *strong preference whole-number-based - strong preference digit-based (5-point scale)*; $M = 3.0$; $r(17) = 0.06$, $p = 0.798$.
- To what extent do you as a teacher prefer an algorithmic over a non-algorithmic approach? *strong preference algorithmic - strong preference non-algorithmic (5-point scale)*; $M = 2.2$; **$r(17) = -0.46$, $p = 0.048$** .
- Which ability do you find more important in general for your students? *performing calculations well on paper - performing calculations well mentally (5-point scale)*; $M = 3.0$; $r(17) = -0.19$, $p = 0.444$.
- How often do you instruct your students in writing down intermediate steps or calculations? *almost never - daily (5-point-scale)*; $M = 4.2$; $r(17) = 0.297$, $p = 0.217$.
- What is more important to you when your students solve multidigit division problems? (*six 5-point scales*).
 - that they write down all calculations - that they try to do it mentally*: $M = 2.4$; $r(17) = -0.26$, $p = 0.275$.
 - that they keep trying until they get the correct solution, even if that takes a lot of time - that they can do it quickly, even if they sometimes make mistake*: $M = 2.5$; $r(17) = -0.29$, $p = 0.234$.
 - that they can determine the exact answer - that they can make a good estimation of the answer*: $M = 3.5$; $r(17) = -0.10$, $p = 0.695$.
 - that they know one solution procedure - that they know multiple solution procedures*: $M = 3.4$; $r(17) = 0.35$, $p = 0.14$.
 - that they use an algorithm - that they choose their own solution strategy*: $M = 3.0$; $r(19) = -0.10$, $p = 0.687$.
 - that they use a method that can always be applied - that they use convenient shortcut strategies (such as $1089 \div 11 = 1100 \div 11 - 1$)*: $M = 3.3$; $r(17) = -0.24$, $p = 0.320$.



The Open Algorithm Based on Numbers (ABN) Method: An Effective Instructional Approach to Domain-Specific Precursors of Arithmetic Development

Gamal Cerda^{1*}, Estíbaliz Aragón², Carlos Pérez³, José I. Navarro² and Manuel Aguilar²

¹ Facultad de Educación, Universidad de Concepción, Concepción, Chile, ² Departamento de Psicología, Universidad de Cádiz, Cádiz, Spain, ³ Dirección de Pregrado, Universidad de O'Higgins, Rancagua, Chile

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

Annemie Desoete,
Ghent University, Belgium
Robert Reeve,
The University of Melbourne, Australia
Natividad Adamuz-Povedano,
Universidad de Córdoba, Spain

*Correspondence:

Gamal Cerda
gamal.cerda@udec.cl

Specialty section:

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

Received: 05 February 2018

Accepted: 06 September 2018

Published: 25 September 2018

Citation:

Cerda G, Aragón E, Pérez C,
Navarro JI and Aguilar M (2018) The
Open Algorithm Based on Numbers
(ABN) Method: An Effective
Instructional Approach
to Domain-Specific Precursors
of Arithmetic Development.
Front. Psychol. 9:1811.
doi: 10.3389/fpsyg.2018.01811

This article presents the results of a comparative study regarding the impact and contribution of two instructional approaches to formal and informal mathematical reasoning with two groups of Spanish students, aged four and five. Data indicated that for both age groups, children under the ABN method [Open Algorithm Based on Numbers (ABN)] ($n = 147$) achieved better results than the group under the CBC approach (Closed Algorithms Based on Ciphers) ($n = 82$), which is the widespread approach in Spanish schools to teach formal and informal mathematical reasoning. Furthermore, the comparative analyses showed that the effect is higher in the group of students who received more instruction on skills considered domain-specific predictors of later arithmetic performance. Statistically significant differences were found in 9 of the 10 dimensions evaluated by TEMA-3 ($p < 0.01$), as well as on estimation tasks in the number-line for the 5-year-old-group. However, the 4-year-old group only presented significant results in calculation and concepts tasks about informal mathematical reasoning. We discuss that these differences arise by differential exposure to specific number-sense tasks, since the groups proved to be equivalent in terms of receptive vocabulary, processing speed, and working memory. The educational consequences of these results were also analyzed.

Keywords: domain-specific precursors, TEMA-3, early arithmetic, ABN method, mathematics

INTRODUCTION

During the 1st years of their lives, students pay special attention to their environment and innately show curiosity about the quantitative relationships that occur around them, thus developing informal mathematical reasoning. These skills are the basis for the mathematical concepts taught at school. As students begin receiving formal instruction, mathematical reasoning is developed and refined (Ginsburg et al., 1998). Children leave aside intuition and develop different types of arithmetic reasoning, such as algebraic reasoning and verbal reasoning, among others. Formal mathematical reasoning requires from students a competent level in the management of symbols and language (Godino and Font, 2003). In particular, formal mathematical reasoning involves conventional knowledge related to number literacy as well as knowledge about the basic concepts

of the decimal system. Furthermore, it also includes the knowledge of number facts and calculation, since instruction and memorization are necessary both for the recovery of number facts as well as for carrying out complex arithmetic operations, such as addition and subtraction with regrouping, solving operations with numbers with middle-zeros, etc. (Robinson et al., 2018).

In order students attain the necessary arithmetic skills for the curriculum content acquisition, it is necessary to ensure they develop from an early age, and in an up-close and meaningful way, contents such as counting as well as relational aspects and processes: problem-solving and number representation (Ginsburg and Baroody, 2007; Alsina, 2012). This is the reason why researchers are focusing their efforts on the study of the mechanisms underlying the arithmetic skills development. One of the main goals is the identification and analysis of the predictor variables for arithmetic performance. They are variables on which the most complex mathematic learning is developed (Cargnelutti et al., 2017; Cerdeja et al., 2015).

There is a long researching studies that defends the role of number sense as a strong predictor of successful mathematics performance, above other general factors such as vocabulary and working memory (Jordan et al., 2007, 2008, 2009). Jordan et al. (2008) define the *number sense* as the ability to understand numbers and arithmetic operations, together with the capacity to make arithmetical judgments resulting from the understanding of numbers and arithmetic facts. According to Jordan et al. (2008), number sense includes the ability to count, knowledge of numbers, and number facts. The topic linking the role of number sense and performance in mathematics is also supported because in having a weak number sense, makes the formal instruction process more difficult for students. This difficulty continues when students go along the compulsory schooling (Baroody and Rosu, 2006). In addition, there is also substantial evidence that accounts the significant relationship between cardinality and number series understanding, achievement of multiplication solving problems, addition and subtraction (De Smedt et al., 2013; Lyons et al., 2014; Vogel et al., 2015).

Counting could also be an important predictor of mathematical performance (Geary, 2011; Bartelet et al., 2014). So, number knowledge involves recognizing differences between quantities and making judgments about the identified quantities. Although younger children rely on visual perception instead of counting to make these type of judgments, however, it has been found that at the age of six, most children incorporate notions of quantities and counting schemes into a mental number-line (Siegler and Booth, 2004). Children associate that numbers presented after other numbers in a counting sequence and, therefore, in a number-line, correspond to a higher number than those presented before. Consequently, students develop a linear representation of numerical magnitudes, which support the learning of the positional value of numbers and the elaboration of mental calculation. This linear representation, along with the operation of counting, should predice mathematics learning difficulties (MLD) (Geary et al., 2009). Cirino (2011) also identified the symbolic and non-symbolic comparison tasks, and the principles or concepts related to competent counting, such as symbolic labeling and

knowledge of the number sequence, as latent variables related to number knowledge.

It has also been maintained that infants can make computation using physical references (Jordan et al., 2008). However, as children start formal education, they begin to use algorithms, understood as sequences of unambiguous instructions used to obtain a required result (Jordan et al., 2008). These sequences contribute to understand some basic mathematical concepts, such as seriations, patterns, forms, comparisons, estimations, verifications, etc. (Levitin, 2015). Key factors associated with the number concept emerge as precursors of mathematical development, and improve the understanding of fundamental arithmetic concepts and facts that, with appropriated instructional approaches, should enable children a higher academic achievement.

In general, teaching methods in formal education are facilitated by the use of textbooks, designed to meet the curriculum needs. These books offer an organized scheme for teaching and learning (Fan et al., 2013; Hadar, 2017). Within the Spanish and international context of mathematics formal instruction, (even if there are differences attributable to teacher's management, textbooks, and schools' pedagogical guidelines), there is a certain homogeneity in terms of teaching methods. Namely, the most extended mathematics teaching approach focuses on learning of additive and multiplicative structures through algorithmic processes based on number figures (Barba-Uriach and Calvo, 2012). We will refer to this approach as the Closed Algorithms Based on Ciphers (CBC) methods.

The algorithms from the CBC methods develop after understanding the numbers' place value. CBC is implemented with numbers. The facts numbers, sequence of steps, and calculations required by the process (which determine the arithmetic tasks), are completely predefined when the quantities involved in the algorithmic are established. In this sense, if two individuals use the same algorithm to perform an arithmetic task, both subjects will necessarily follow the same intermediate steps. Moreover, in a traditional CBC approach, the partial calculations completed is not a requirement to the problem that is being solved. Thus, they become only crucial steps for the correct functioning of the algorithm. In addition, in a CBC approach, there is more room for error, because non-specific reference mechanisms to monitor the problem's partial steps are not trained.

Faced with this condition, several Spanish countries school systems are currently implementing the Open Algorithm Based on Numbers (ABN) method (Martínez-Montero and Sánchez, 2013). This approach gathers and addresses the natural route of each stages of the exploratory process that children utilize to understand numbers and their properties. This route is used as the background to design the instructional sequences and significant algorithms to perform facts with numbers. The ABN method has precedents in some educational proposals launched in the Netherlands to renew the mathematics teaching and learning in general and particularly teaching methodology for calculation. This was called "realistic mathematics." This is oriented to development mathematical competence and fostering mathematical reasoning through manipulative and stimulating

instruments for students in order to increase motivation and attention.

Implementation of ABN procedure begins in the academic course of 2008/2009 in a group of first grade of primary school in a public school at Cadiz (Spain). One year later, ABN extends in 4 more schools in the same province with approximate 125 students first graders. During the 2011–2012 different nationwide schools started implementing ABN method, distributed by more than 10 Spanish autonomous communities. Throughout 2013 keep growing the Spanish public schools using this method, and starting from the 1st year of preprimary education. In addition ABN begins to expand internationally in other countries such as Mexico, Argentina or Chile. But there are still no published results on these experiences. According to data provided by Cantos (2016) currently between 6,000 and 7,000 classrooms follow the ABN methodology, representing an approximate total number of 200,000 students learning math with ABN.

The integration of this path to the ABN method also provides an adequate understanding of the different steps for calculation process. Thus, with regard to the additive and multiplicative algorithmic, these processes can be understood as a transformation of resulting quantities from a dynamic operative process. Namely, a process in which each sequence is a sub-process involving explicit quantities, which are operated under the user's criteria. Starting in early childhood, the ABN method focuses on performing math tasks with whole numbers, according to ranges or universes numbers. The method uses different formats for representation and manipulation of specific numbers, according to concrete, pictorial, abstract (CPA) instructional approach to teaching, that develops a deep and sustainable understanding of maths in pupils (Bruner, 1966). The principle of the ABN method is to maintain the numerosity of quantities all the times, in terms of knowledge, composition and decomposition, as well as taking into consideration how they operate in relation to other quantities (Martínez-Montero, 2010, 2011). In recent years, there has been evidence of the effect that this new methodology should have: to learn mathematics in a realistic and quotidian way, moving away from the mere acquisition of strategies and knowledge of symbols, that does not guarantee a realistic understanding (Bracho-López et al., 2014a). In this way, the unidirectional and sequential transfer of information is avoided, to give way to a more active learning process, in accordance with the context of student development (Novo et al., 2017).

With ABN approach, algorithmic are latent parts in the process of structuring mathematical knowledge. A meaningful mathematical understanding is developed throughout schooling. This knowledge structure also contributes to early approaches to Algebra, by experimentally incorporating mathematical concepts such as equations, powers, roots, etc. ABN approach requires focusing on the early skills acquisition to allow adequate understanding of the number but not only to reel number sequences off by memory. Thus, the ABN method focuses on developing in children the ability to establish differences or similarities between groups; relations between objects by grouping them according to specific criteria; pairing set elements

with only one element from a different set; to intuit the order of objects according to number ranges; and, to use acquired problem-solving skills to elucidate daily life problems that involve counting. The flexibility that characterizes the ABN method not only offers an advantage in the development of original solution approaches or different types of solutions, but it also provides a set of strategies to adequately solve mathematic tasks (Torbeys et al., 2005).

This previous background allows us to claim that a timely and adequate assessment of informal and formal reasoning (both regarding skills and concepts associated with early mathematical competence), can significantly contribute to the analysis and prediction of students achieving. Both dimensions are precursors of students' performance in mathematics. In the same way, given the emphasis of the ABN method in working with quantities from the early stages, it is relevant to examine whether or not the exposure to a specific instructional approach for the development of mathematics skills, concepts and principles in early childhood education produces a differential level of development in students. In short, this research aims to determine potential differences linked to a specific instructional approach to formal and informal mathematical reasoning in a group of preschool children.

The following three hypotheses are proposed for this study:

- (1) There are significant differences for ABN method students group in comparison to CBC group, in informal mathematical reasoning. Specifically in counting, comparisons, calculation and informal concepts, regardless of the school year.
- (2) There are significant differences for ABN method students group in comparison to CBC group, in the formal mathematical reasoning. Specifically in number facts, formal calculation, and formal mathematical concepts, regardless of the school year.
- (3) There are significant differences for ABN method students group in comparison to CBC group, in number-line estimation, regardless of the school year.

MATERIALS AND METHODS

A quasi-experimental descriptive and comparative cross-sectional design was used. Measurements of the dependent variables were taken in a single moment. Mathematical performance was compared in two different ages groups of students (4 and 5 years), with two types of mathematical instruction approaches (ABN and CBC). Considering the characteristics of the cross-sectional designs, three measures of control over cognitive variables were used in order to guarantee the groups' equivalence and comparability. The designs of causal type, with control variables such as those used in this work, allow to establish some relevant inferences, given that the cause-effect relationships already occurred or occur during the process of measuring (Hernández et al., 2014). These kinds of designs have been used for that purpose in other similar qualified investigations (Wang et al., 2016).

Participants

The total sample of students ($n = 224$) belonged to five public schools from Spain, four of them located in the community of Andalusia and one of them in the community of Madrid. In the Spanish public school system, most of the parent's choice the school according with the standard of proximity to their homes. This criterion maintains the patterns of social stratification associated with residential zone (Mancebon-Torrubia and Perez Ximenez-de-Embun, 2014). Social and economic neighborhood similarities are usually coincident with the social and economic structure of schools settled on the same neighborhood. According to the OECD (2015) report, Spanish school system holds one of the highest social-economical homogeneity rates in classrooms. In this study, schools were settled in middle-class neighborhoods, considering that social-economic differences between students were not relevant. Teachers' background was also similar, considering that in the Spanish education system all Pre-school and Primary school teachers hold a University degree. Thus, the socioeconomic composition of these five public schools corresponds to middle-class standards. The participants were 224 students, of which 110 (49.1%) girls and 114 boys (50.9%). The average age for the female group was of 65.47 months ($SD = 6.98$). The average age for the male group was of 64.33 months ($SD = 6.45$). 111 students belonged to the 4-year-old pre-school class, and 113 to the 5-year-old pre-school class. Although so far ABN is being used from 4 to 14 years old students, we consider that the beginning of the schooling and the first contact for children with the formal academic math is important for learning this subject. Several longitudinal studies shown that when young children start having trouble with mathematic, they keep on this problem later (Navarro et al., 2012).

Students taught under the Open Algorithm Based on Numbers method (ABN group) were 142; 74 aged 4, and 68 aged 5. Student under the Closed Algorithms Based on Ciphers (CBC group) were 82; 37 aged 4 and 45 aged 5. Students with special educational needs were not included in this study.

Procedure

Trained professionals carried out participants' assessment in two sessions. Each was of approximately 15–20 min. The purpose of this design was to attend to the particular characteristics of the students, and to avoid student's tiredness. The evaluation conditions were optimal. The assessment was conducted in settings free of distractions that could interfere in the results. One session consisted in the administration of TEMA-3 test to evaluate students' mathematical competence. The other session focused on the evaluation of the control variables (verbal working memory, receptive vocabulary, and processing speed). During this session, the number-line estimation test was also administered. Harvey and Miller (2017) reported that receptive vocabulary significantly affects to early math skills. Also, Peng et al. (2016) considered that processing speed and working memory as variables related to mathematical competence. These two subtests were used as control tests to establish the equivalence of the groups.

This study considered two groups of students, which will be referred to CBC-group and ABN-group. The CBC-group consisted of students who received instruction under the CBC approach. This type of education is widespread in most Spanish and other countries schools, and it characterizes for adjusting to the contents required by the educational administration of the country. CBC methodology focuses on the monitoring of content learning through textbooks. The didactic proposal of textbooks is mostly oriented to a CBC additive and multiplicative structure. The ABN-group was composed of students who received mathematical instruction through the ABN method. The teachers in charge of this group had a specific training in the ABN instructional method. Teachers' training took into account the contents, competencies and specific goals required by the educational administration government for each grade. Both groups of participants received the compulsory mathematics contents stated in the school curriculum for each grade, but with different approach. Instructional timing was the same for both groups and it was accordingly to the instructional schedule established by the Spanish Ministry of Education. Thus, the significant difference between both groups was the mathematics instructional approach used. It is important to note that all participating students received mathematics instruction through one method or the other from the 1st year of preschool education.

All subjects gave written informed consent in accordance with the Declaration of Helsinki. Informed consent was obtained from parents, teachers and school principals involved in this study.

Instruments

Test of Early Mathematics Ability-Third Edition, TEMA-3 (Ginsburg and Baroody, 2007)

This test assesses mathematical competence and consists of two subtests that focus on the evaluation of informal and formal reasoning, both in terms of skills and concepts. The informal reasoning subtest is composed of tasks aimed at the assessment of counting, comparison of quantities, informal calculation and basic informal concepts. The formal reasoning subtest evaluates conventions related to number quantity literacy, knowledge of number facts, formal calculation, and formal mathematical concepts.

TEMA-3's administration was individual taking around 30 min. Administration timing differs according to the student's age. The test is applicable to children aged between 3 and 9 years old. TEMA-3 has 72 items presented in order of increasing difficulty. The Cronbach's alpha for this test was 0.91 for 4-year-olds and 0.95 for 5-year-olds.

In the Spanish standardization for TEMA-3, it is reported the following scores: 4 years ($M = 15.25$, $SD = 5.89$), and 5 years ($M = 25.03$, $SD = 7.23$). The range data (minimum and maximum score) do not appear in the Spanish standardization manual.

Numerical Estimation Task (Siegler and Booth, 2004)

This pencil-and-paper test evaluates estimation skills in a number-line. For its administration, participants are presented with a sheet of paper with a 20-centimeter number line, which starts at zero and ends at 20. Above the line, in the upper central part of the sheet, a number is shown. Participant must point out

the number in the straight line. The test consists of 10 items, which correspond to the following numbers: 2, 4, 7, 8, 11, 13, 16, 17, 18, and 19, randomly presented. The mean comparison rate was calculated according to the number of correct answers with respect to the number requested by the test versus the number provided by the student. Answers were considered correct if they did not present a rate of error higher than $\pm 15\%$ for the requested number. The Cronbach's alpha for this test was 0.80

Coding Subtest From the Wechsler Preschool and Primary Scale of Intelligence, Third Edition (WPPSI-III) (Wechsler, 2009)

This test is included within the Wechsler Intelligence Scale for preschool and primary school (Wechsler, 2009). It assesses processing speed, visual perception, visual-manual coordination, short-term memory, learning ability, and cognitive flexibility. The student must complete a set of 64 figures presented with the appropriate symbols. Participant must follow the reference models within a time limit of 2 min. The Cronbach's alpha for this test was 0.84.

Receptive Vocabulary Test From the Dyslexia Screening Test - Junior (DST-J) (Fawcett and Nicolson, 2013)

This test is a measure of vocabulary mastery and reasoning ability. The purpose of this test is to evaluate receptive vocabulary through a multiple-choice format. The test comprises 18 items; each correct item receives one point. The Cronbach's alpha for this test was 0.74.

Backward Digit Task From the Dyslexia Screening Test - Junior (DST-J) (Fawcett and Nicolson, 2013)

This test measures verbal working memory. It involves the oral repetition of digits in reverse order. As the number of trials increases, the number of digits increases and, consequently, the difficulty of the task. This task is composed of seven series of

two items each. The test includes three items plus two additional, administrated in case that child has difficulties in properly understanding the instructions. The Cronbach's alpha for this test was 0.85

Statistical Analyses

In order to calculate the comparative analyzes between average scores obtained by the ABN and CBC groups, one-way ANOVA tests were completed. Whenever the homoscedasticity of the variances was not proven, a correction of the degrees of freedom, and Welch's robust test was applied. The effect size was also calculated for the total variables measured.

RESULTS

In order to establish that the ABN and CBC groups were equivalent, three control tests were computed. A receptive vocabulary test, a processing speed test, and a working memory test of backward digits.

Table 1 shows the correlations matrix of the scores and total scores of the subtests of formal and informal mathematical thinking of the TEMA-3 test reached by the students, with the purpose of analyzing the intensity of the associations between them. It is observed that all of them are statistically significant.

In order to examine if there were differences in the scores of these variables according to the age group, comparisons were made by means of simple ANOVA tests.

No significant differences were found in the receptive vocabulary test results for the 4-year-old group [$Mdn_{CBC} = 11.56$, $SD_{CBC} = 1.96$; $Mdn_{ABN} = 12.17$, $SD_{ABN} = 1.82$; $F(1,109) = 2.604$, $p > 0.01$]. Likewise, no significant differences were found for the WPPSI [$Mdn_{CBC} = 27.02$, $SD_{CBC} = 11.20$; $Mdn_{ABN} = 26.79$, $SD_{ABN} = 11.03$; $F(1,109) = 0.011$, $p > 0.01$] and neither were there significant differences in the comparison between the backward digit test for the 4-year-old group [$Mdn_{CBC} = 1.48$,

TABLE 1 | Correlation matrix of the student scores in formal and informal mathematical thinking subtest of TEMA-3.

	Informal reasoning	Counting	Comparing	Informal calculations	Informal concepts	Formal reasoning	Conventions	Number facts	Formal calculations	Concepts
Informal reasoning	1	0.967**	0.793**	0.854**	0.658**	0.853**	0.856**	0.476**	0.277**	0.442**
Counting		1	0.677**	0.747**	0.573**	0.860**	0.876**	0.470**	0.263**	0.402**
Comparing			1	0.662**	0.466**	0.638**	0.643**	0.323**	0.218**	0.376**
Informal calculations				1	0.537**	0.676**	0.680**	0.354**	0.198**	0.403**
Informal concepts					1	0.546**	0.464**	0.472**	0.309**	0.345**
Formal reasoning						1	0.944**	0.692**	0.476**	0.489**
Conventions							1	0.452**	0.261**	0.367**
Number facts								1	0.622**	0.253**
Formal calculations									1	0.192**
Concepts										1

** $p < 0.01$.

$SD_{CBC} = 1.30$; $Mdn_{ABN} = 1.63$, $SD_{ABN} = 1.47$; $F(1,109) = 0.251$, $p > 0.01$].

Similarly, no significant differences were found in the receptive vocabulary test results for the 5-year-old group [$Mdn_{CBC} = 12.57$, $SD_{CBC} = 1.58$; $Mdn_{ABN} = 12.67$, $SD_{ABN} = 1.38$; $F(1,111) = 0.122$, $p > 0.01$]. Likewise, no significant differences were found in the results for the WPPSI [$Mdn_{CBC} = 34.80$, $SD_{CBC} = 9.43$; $Mdn_{ABN} = 35.64$, $SD_{ABN} = 10.43$; $F(1,111) = 0.192$, $p > 0.01$]; and neither were there significant differences in the comparison between the backward digit test for the 5-year-old group [$Mdn_{CBC} = 2.22$, $SD_{CBC} = 1.44$; $Mdn_{ABN} = 2.91$, $SD_{ABN} = 1.33$; $F(1,111) = 6.765$, $p > 0.01$].

Several comparative analyzes were carried out in order to guarantee equivalence and comparability among ABN and CBC groups (gender, mathematical instructional method and autonomous community): (a) *Gender (4-year-old)*. For the 4 years aged group, vocabulary [$t_{(109)} = 0.267$; $p > 0.05$], processing speed [$t_{(109)} = -2.314$; $p > 0.05$], and working memory [$t_{(109)} = -0.336$; $p > 0.05$] differences between groups were not significant for gender; (b) *Mathematical instructional method (4-year-old)*. In the same way, there were no statistically significant differences in vocabulary [$t_{(109)} = -1.614$; $p > 0.05$], processing speed [$t_{(109)} = 0.103$; $p > 0.05$], and working memory [$t_{(109)} = -0.501$; $p > 0.05$] for math instructional method used; (c) *Autonomous community (4-year-old)*. Vocabulary [$t_{(109)} = -0.475$; $p > 0.05$], processing speed [$t_{(109)} = 1.784$; $p > 0.05$], and working memory [$t_{(109)} = 0.351$; $p > 0.05$] differences between groups were not significant for autonomous community of the schools where students attended; (d) *Gender (5-year-old)*. For the 5 years aged group, vocabulary [$t_{(109)} = 0.267$; $p > 0.05$], processing speed [$t_{(109)} = -2.314$; $p > 0.05$], and working memory [$t_{(109)} = -0.336$; $p > 0.05$] differences between groups were not significant for gender; (e) *Mathematical instructional method (5-year-old)*. In the same way, there were not statistically significant differences in vocabulary [$t_{(109)} = -1.614$; $p > 0.05$], processing speed [$t_{(109)} = 0.103$; $p > 0.05$], and working memory [$t_{(109)} = -0.501$; $p > 0.05$] for method of instruction; (f) *Autonomous community (5-year-old)*. Either, in vocabulary [$t_{(109)} = -0.475$; $p > 0.05$], processing speed [$t_{(109)} = 1.784$; $p > 0.05$], and working memory [$t_{(109)} = 0.351$; $p > 0.05$] for autonomous community of the schools where students attended, no statistically significant differences were found.

Table 2 shows TEMA-3 subtests scores comparing the two instructional methods (CBC and ABN).

Statistically significant differences were found ($p < 0.05$) between the CBC and ABN groups in informal calculation and informal concepts dimensions, although effect sizes were small. In formal reasoning calculations dimension, 4-year-old children were not able to correctly solve any task. A possible explanation should be, because the suspension criterion was used before being able to solve them. The same works to number facts dimension, where 4-year-old students did not solve any tasks. However, some ABN-group participants appropriately solved up to two tasks of this type (**Table 3**).

Since five different schools participated in this study, a cross-school analysis has been carried out. This statistical analysis generated two categories of schools according to the Autonomous

Community to which they belong: Andalusia and Madrid. The purpose was to explore whether there were differences. Comparing the average scores in informal and formal thinking, in each age group analyzed, independently of the instructional method, few statistically significant differences in most of the dimensions explored were found. For 4 year-old groups no statistically differences were found in informal reasoning (counting, comparing, informal calculations, informal concepts), either formal reasoning (conventions, number facts, formal concepts, comparing). For the 5 year-old group no statistically significant differences were found in comparing, number facts and formal concepts. However, differences were found for 5 year-old group in informal reasoning [$F_{(1,111)} = 19.011$, $p < 0.05$, $\eta^2 = 0.146$], and the following subtests of this component: counting [$F_{(1,111)} = 18.249$, $p < 0.05$, $\eta^2 = 0.141$]; informal calculations [$F_{(1,111)} = 13.616$, $p < 0.05$, $\eta^2 = 0.109$], and informal concepts [$F_{(1,111)} = 20.331$, $p < 0.05$, $\eta^2 = 0.155$]. Similarly, differences were found in formal reasoning [$F_{(1,111)} = 14.67$, $p < 0.05$, $\eta^2 = 0.117$]: conventions [$F_{(1,111)} = 13.565$, $p > 0.05$, $\eta^2 = 0.109$], and formal calculations [$F_{(1,111)} = 1.954$, $p > 0.05$, $\eta^2 = 0.108$].

Statistically significant differences were found in 9 out of 10 dimensions compared ($p < 0.01$). In particular, counting dimension, which is part of the development of informal reasoning, a large effect size was found. In addition, significant differences and a large effect size were found of convention dimensions, which is part of development of formal reasoning.

In order to support that instructional approach generated a positive interaction cross-age effect, an additional statistical analyzes were carried out. This effect reproduces time effect learning with the instruction methodology (ABN or CBC), by observing the informal and formal mathematical thinking tasks data analyzed.

A significant disordinal interaction between the teaching method and students age was found, regarding the level of mathematical informal thinking: The methods' effect was not the same for each age students group, but the difference was always for the ABN group [$F_{(1,220)} = 10.68$; $p < 0.05$, $\eta^2 = 0.046$]. Similarly, statistical differences were found considering the learning method [$F_{(1,220)} = 37.11$; $p < 0.05$, $\eta^2 = 0.144$]; and by age group [$F_{(1,220)} = 60.34$; $p < 0.05$, $\eta^2 = 0.215$]. These main effects indicated that ABN method' students achieved higher than CBC method' students. Furthermore, 5-year-old students group achieved better in math informal thinking than their 4 years peers; (a) *Mathematical formal thinking dimension comparison*. Regarding the math formal thinking dimension, a significant interaction effect between instructional method and age group was found [$F_{(1,220)} = 16.39$; $p < 0.05$, $\eta^2 = 0.069$]. In the same way, differences were found by learning methodology (ABN or CBC) [$F_{(1,220)} = 28.94$; $p < 0.05$, $\eta^2 = 0.116$]; and by age group [$F_{(1,220)} = 40.01$; $p < 0.05$, $\eta^2 = 0.154$]. Concerning the dimensions that conform the math formal thinking, interaction effect was found in conventionalism [$F_{(1,220)} = 17.99$; $p < 0.05$, $\eta^2 = 0.076$]; but none in numerical facts [$F_{(1,220)} = 3.09$; $p > 0.05$, $\eta^2 = 0.014$]. Main effects were found according to the ABN or CBC method [$F_{(1,220)} = 6.96$; $p < 0.05$, $\eta^2 = 0.031$]; and by age group [$F_{(1,220)} = 8.50$; $p < 0.05$, $\eta^2 = 0.037$]. None interaction

TABLE 2 | Descriptive statistics for the different TEMA-3 test dimensions for the 4-year-old CBC and ABN groups.

Dimensions	Group	N	M	SD	CI 95%	MIN	MAX	F	η^2
Informal reasoning	CBC	37	17.35	4.87	15.72–18.98	8	27	3.428	0.030
	ABN	79	19.18	6.03	17.82–20.53	4	35		
	Total	116	18.59	5.73	17.54–19.65	4	35		
Counting	CBC	37	11.22	2.95	10.23–12.20	6	19	2.392	0.021
	ABN	79	12.16	3.67	11.34–12.99	2	22		
	Total	116	11.86	3.47	11.22–12.50	2	22		
Comparison	CBC	37	2.51	1.07	2.15–2.87	1	5	0.486	0.004
	ABN	79	2.62	1.19	2.35–2.89	1	5		
	Total	116	2.59	1.15	2.37–2.80	1	5		
Informal calculation	CBC	37	2.03	1.30	1.59–2.46	0	4	4.154*	0.037
	ABN	79	2.52	1.35	2.21–2.82	0	5		
	Total	116	2.36	1.35	2.11–2.61	0	5		
Informal concepts	CBC	37	1.59	0.60	1.39–1.79	0	2	4.885*	0.043
	ABN	79	1.81	0.60	1.67–1.94	0	3		
	Total	116	1.74	0.61	1.62–1.85	0	3		
Formal reasoning	CBC	37	3.00	1.22	2.59–3.41	1	6	1.302 (<i>FWelch</i>)	0.009
	ABN	79	3.30	1.88	2.88–3.72	0	10		
	Total	116	3.21	1.70	2.89–3.52	0	10		
Conventions	CBC	37	2.08	1.14	1.70–2.46	0	5	1.345 (<i>FWelch</i>)	0.010
	ABN	79	2.35	1.52	2.01–2.69	0	6		
	Total	116	2.27	1.41	2.00–2.53	0	6		
Number facts	CBC	37	0	0	0	0	0	1.842	0.017
	ABN	79	0.07	0.30	0.0–0.14	0	2		
	Total	116	0.05	0.25	0.0–0.09	0	2		
Formal calculation	CBC	37	0	0	0	0	0		
	ABN	79	0	0	0	0	0		
	Total	116	0	0	0	0	0		
Concepts	CBC	37	0.92	0.28	0.83–1.01	0	1	0.165	0.002
	ABN	79	0.89	0.35	0.82–0.97	0	2		
	Total	116	0.90	0.33	0.84–0.96	0	2		
TEMA-3 Total score	CBC	37	20.35	5.84	18.40–22.30	11	33	2.968	0.027
	ABN	79	22.48	7.60	20.77–24.18	4	45		
	Total	116	21.80	7.13	20.49–23.11	4	45		

* $p < 0.05$.

effects were found in formal calculus [$F_{(1,220)} = 1.46$; $p > 0.05$, $\eta^2 = 0.007$]; although the main effect attributable to the age group was found [$F_{(1,220)} = 4.10$; $p < 0.05$, $\eta^2 = 0.018$], but not by method [$F_{(1,220)} = 1.46$; $p > 0.05$, $\eta^2 = 0.007$], neither in formal mathematical thinking [$F_{(1,220)} = 2.58$; $p > 0.05$, $\eta^2 = 0.012$]. A main effect attributable to the age group was observed [$F_{(1,220)} = 3.94$; $p < 0.05$, $\eta^2 = 0.018$], rather than by the method [$F_{(1,220)} = 0.81$; $p > 0.05$, $\eta^2 = 0.004$]; (b) *TEMA-3 score comparison*. Regarding the TEMA 3 direct scores, a significant interaction between mathematics learning method (ABN or CBC) and the students' age [$F_{(1,220)} = 13.05$; $p < 0.05$, $\eta^2 = 0.056$] was found between. Significant differences were also observed by the learning method [$F_{(1,220)} = 37.92$; $p < 0.05$, $\eta^2 = 0.147$]; and by age group [$F_{(1,220)} = 59.44$; $p < 0.05$, $\eta^2 = 0.213$]. These main effects suggest that participants who learned mathematics with the ABN approach achieved higher results than students who learned with the CBC method. Similarly, 5-year-old students achieved higher average scores in informal mathematical thinking than those in 4-year-old group.

For numerical estimation tasks, statistically significant differences were found in 5-year-old group. Specifically, when comparing student averages regarding the ability to estimate numbers, significant differences were found for the ABN-group, with a small effect size (Table 4).

DISCUSSION

The previous results support the hypothesis about the positive impact of the ABN method on the dimensions that make up formal and informal reasoning. Namely, the participants who were under the ABN approach obtained significantly higher results than the CBC approach. Thus, it can be maintained that the ABN method provides an integrated perspective, based on the significant learning of the decimal counting system, as well as providing complete understanding of basic math processes and their properties (Martínez-Montero, 2010, 2011; Martínez-Montero and Sánchez, 2011). Even so, results indicate that

TABLE 3 | ANOVA and effect size results of TEMA-3 dimensions for the 5-year-old CBC and ABN groups.

Dimensions	Group	N	M	CI 95%	MIN	MAX	F	η^2
Informal reasoning	CBC	45	20.69	19.17–22.21	5	32	51.58**	0.317
	ABN	68	27.63	26.41–28.85	14	36		
	Total	113	24.87	23.73–26.00	5	36		
Counting	CBC	45	12.71	11.71–13.71	4	21	51.02**	0.315
	ABN	68	17.40	16.55–18.24	9	22		
	Total	113	15.53	14.76–16.30	4	22		
Comparing	CBC	45	3.22	2.92–3.52	1	5	15.25**	0.121
	ABN	68	3.99	3.73–4.23	1	6		
	Total	113	3.68	3.48–3.88	1	6		
Informal calculations	CBC	45	2.98	2.60–3.35	0	5	21.33**	0.161
	ABN	68	3.94	3.70–4.18	1	5		
	Total	113	3.56	3.33–3.78	0	5		
Informal concepts	CBC	45	1.82	1.66–1.98	0	3	19.29**	0.148
	ABN	68	2.29	2.15–2.43	1	4		
	Total	113	2.11	1.99–2.22	0	4		
Formal reasoning	CBC	45	3.58	3.01–4.14	0	12	40.69**	0.268
	ABN	68	5.97	5.48–6.45	2	12		
	Total	113	5.02	4.59–5.44	0	12		
Conventions	CBC	45	2.53	2.12–2.95	0	7	56.74**	0.338
	ABN	68	4.43	4.12–4.73	1	7		
	Total	113	3.67	3.37–3.97	0	7		
Number facts	CBC	0.08	0.46	–0.05–0.22	0	3	7.13** (<i>FWelch</i>)	0.049
	ABN	0.42	0.86	0.21–0.63	0	4		
	Total	0.29	0.75	0.15–0.43	0	4		
Calculations	CBC	0.02	0.14	–0.02–0.06	0	1	2.04 (<i>FWelch</i>)	0.014
	ABN	0.08	0.33	0.00–0.16	0	2		
	Total	0.06	0.27	0.01–0.11	0	2		
Concepts	CBC	0.93	0.25	0.85–1.00	0	1	5.85*	0.050
	ABN	1.02	0.17	0.98–1.07	1	2		
	Total	0.99	0.21	0.01–0.95	0	1		
TEMA-3 Total score	CBC	45	24.27	22.23–26.30	5	43	52.26**	0.320
	ABN	68	33.60	31.98–35.22	17	47		
	Total	113	29.89	28.37–31.40	5	47		

** $p < 0.01$, * $p < 0.05$.**TABLE 4 |** Descriptive and inferential analyses results for the numerical estimation task for CBC and ABN groups.

Course	Method						F	η^2
	CBC			ABN				
	M	SD	N	M	SD	N		
4 year-old	2.46	1.80	37	3.22	1.96	74	3.871	0.034
5 year-old	2.58	2.01	45	3.94	1.99	68	12.50**	0.101
Total	2.52	1.91	82	3.56	2.05	142	14.43**	0.061

** $p < 0.01$.

this differential gain is substantial for the 5-year-old group, where statistically significant differences were found in all formal and informal mathematical reasoning dimensions. For the 4-year-old children, informal calculations and informal concepts dimensions were found statistically significant.

A potential explanation for these results is that children in the 5-year-old group have received 2 years of systematic formal instruction in mathematics, which has included both ABN and CBC approaches. Therefore, the ABN group students showed a higher ability to successfully solve the tasks that compose each of the dimensions associated with formal and informal reasoning, assessed by TEMA-3 test, because the test scores provide a standardized measure of early arithmetic performance (Núñez and Lozano, 2009; Ryoo et al., 2015). These results are relevant, because the tested skills for informal mathematical reasoning, such as the ability to pay selective attention to numbers contribute to the development of the number sense. This is one of the main predictors of arithmetic performance. In addition, this type of number acuity reinforces mathematical achievement in early childhood, even though the influence of non-numerical characteristics significantly decreases when children developmental progresses. Even so, the ability to pay selective attention remains as a

determining factor for decision-making at all ages (Starr et al., 2017).

Similarly, the ability to compare properly and quickly numerical magnitudes, was a predictor of achievement in mathematics, independent of age, intellectual capacity, and number identification speed (De Smedt et al., 2009; Fazio et al., 2014). Several studies have found that babies are able to selectively pay attention to numbers and size, which could be considered to be the basis for the development of number-sense (Cantrell and Smith, 2013; Mou and vanMarle, 2014; Szkudlarek and Brannon, 2017). Furthermore, there is evidence that number representations, even in adulthood, are influenced by non-numerical properties, such as the size of stimuli, for example (Defever et al., 2013; Fuhs and McNeil, 2013; Gilmore et al., 2013; Szucs et al., 2013).

Likewise, there is significant evidence of the relationship between processing of cardinality and number seriation and arithmetic achievement. Namely, this is found in conditions involving easy multiplication, addition and subtraction tasks (De Smedt et al., 2013; Lyons et al., 2014; Chu et al., 2015; Vogel et al., 2015). Therefore, to improve all these capacities is optimistic, because it provides a better understanding of the cognitive architecture underlying achievement in early childhood education in mathematics, and the impact of different approaches for teaching and learning (Lo et al., 2017).

In summary, in the case of the 5-year-old group, significant differences were verified for the total number of dimensions that make up informal reasoning. In the case of the 4-year-old group, statistically significant differences for the ABN group were found just in calculation and informal concepts dimensions. Consequently, no significant differences were found for counting and comparison of quantities dimensions for this group. These results can be explained because ABN method produces differences achievement for the 4-year-old group for more complex informal tasks; in which students need not only knowledge of numbers, but also the management and application of resolution strategies.

With respect to the second hypothesis, for the 5-year-old group, the boys and girls under the ABN method showed a better performance on the direct score associated with formal reasoning than their peers under the CBC methodology did. Formal reasoning is a very relevant dimension of children's mathematical reasoning, since it implies knowledge and skills, such as the conventions of reading and writing of quantities, the command of number facts, and formal calculation. In this sense, our findings provided relevance to the accuracy of formal procedures and to the basic concepts of the decimal system, such as space value and equivalences between different orders of magnitude. Even though the ABN method has been recently incorporated in Spanish school curricula, these findings coincide with other research previously conducted about the ABN method (Adamuz-Povedano and Bracho-López, 2014; Bracho-López et al., 2014b; Aragón et al., 2017a,b).

Considering the convention dimensions, the existence of errors in the literacy of the quantity for single items may indicate that certain rules have not been fully learned. This finding becomes clearer when other variables are controlled, such as

the educational level of students' parents and their previous levels of literacy. The ABN method contributes in making these operations automatic. In addition, the results obtained in this research are consistent with a previous study with primary school students. Bracho-López et al. (2014b) showed that a student's following the ABN method got significant differences compared to their peers under the CBC approach in tasks involving the decimal system and number facts. Likewise, Moore et al. (2016) demonstrated that the cardinal knowledge exhibited by preschoolers, as well as their competence in manipulation of quantities associated with symbolic numerals, predicted higher flexibility in processing of magnitudes and in academic performance in the future. In this respect, the ABN group obtained significantly higher results in the domains of number facts and calculation when compared with the CBC group.

The significance of these findings lies in the fact that as children progress in a formal instruction, they are expected to begin using algorithms that synthesize sequences of unambiguous instructions to obtain a required result; algorithms that are also assumed to contribute to the basic understanding of mathematical concepts (Levitin, 2015). Hence, finding a better performance in these tasks in the ABN group evidences the strengthening and earlier consolidation of these type of procedural logic operations of abstract nature. In the same way, it is very important to achieve proficiency of number facts, since they are one of the central goals for the mathematics teaching in early education. Namely, students in their 1st years of formal education must be able to remember and provide fast responses to the calculations involved in basic tasks. Domain number facts facilitates and speeds up these processes, leaving space and opportunity to better understanding. It is worth to mention that a good command of this skill does not only involve memory, but is also linked to the application of previously stored rules, which allows easier access to the processing of symbolic magnitudes (De Smedt et al., 2013).

Finally, with respect to formal calculation, the ABN students also obtained better results than their peers following the CBC methodology. It is important to remember that performing formal calculations entails a command of the decimal number system, counting strategies, and knowledge of number facts. In the case of the ABN method, algorithmic operations are part of learning activities and contribute to meaning making. These two facts contribute to a better predisposition and attitude toward to complete the task, which has a favorable impact in the conceptual knowledge that children incorporate and bring into play when making strategic decisions for the resolution of problems (Robinson and Dubé, 2012).

The third hypothesis of this research focused around of significant differences between both approaches concerning numerical estimation in the number-line tasks. The results showed significant differences for the 5-year-old ABN group. According to a previous study, this advantage was also at the age of 6 years (Aragón et al., 2017b). The importance of this result lies in the fact that estimation skills contribute to a good arithmetic competence, since the proficiency in these tasks involves capacity to provide meaning to the magnitude of numbers in a number-line (Laski and Siegler, 2007). However, numerical estimation

tasks stand out as one of the most complex numerical activities for 5-year-old students (Araújo et al., 2014). As students progress in a formal instruction, they make fewer errors and improve their accuracy (Siegler and Opfer, 2003). Therefore, an adequate training in basic concepts underlying numerical estimation skills (as the ABN method does), contributes to improve this ability. Estimation is a relevant part of the prerequisites of early arithmetic skills and contributes to a good performance in mathematics in later stages of formal instruction. Concerning this matter, Watts et al. (2014) evaluated early arithmetic variables in early childhood education and confirmed that early arithmetic competences were predictors for achievement in primary, and also secondary education.

In short, the impact of the mathematical structure developed by the ABN method leads to a better understanding of several essential operations of Algebra (roots, powers, functions, for example). This is because algebra is a discipline in close conceptual relationship with arithmetic. However, in the case of Algebra, the processes and concepts involve a bigger capacity for abstraction. This capacity could be promoted through higher skills for arithmetic representation and could lay the foundations for algebraic representation (Humberstone and Reeve, 2018). Furthermore, this relationship could also explain the higher levels of achievement exhibited by the ABN children in formal and informal reasoning dimensions assessed.

Finally, it should be considered that the differences between the ABN and CBC methodologies are not limited to the mere use of a specific type of algorithmic strategy. The differences between the two approaches lies in understanding the mathematical structure underlying arithmetic operations, which is a critical factor for the development of many others mathematical skills (National Governors Association Center for Best Practices and Council of Chief State School Officers, 2010). From this point of view, we believe that the ABN method defines the processes for arithmetic operations by using methods of decomposition and tasks with quantities. It allows a better understanding of the underlying mathematical structure, as well as better predisposition and basis for the understanding of problems of an additive and multiplicative nature. It is well established that additive and multiplicative tasks share some concepts such as identity, negation, commutativity, equivalence, reversal, and associativity (Robinson et al., 2018). They are key for building the formal arithmetic knowledge in the algebra constructing phases.

Study Limitations and Future Perspectives

Since this research was not an experimental time-series design, we are not able to attribute conclusively the difference in formal and informal mathematical reasoning to the instructional approaches used. Even so, the control tests used established that the qualifications of the ABN and CBC groups compared equivalent. Therefore, we can infer that the differences found can be function of the instructional approach, given that the immersion time of both groups corresponds to the same number of school years.

It should also be mentioned that the effect of the teacher instruction, in terms of years of service, initial teacher training approaches, and/or gender, as well as the students themselves, are also limitations that could have had an effect on the differential results observed, in the sense of mitigating or enhancing the observed differences.

Therefore, for future research and in order to have control over the sources challenging internal and external validity, a quasi-experimental longitudinal study is proposed. We believe that this type of design will strengthen the hypothesis about the favorable impact of the ABN instructional approach on informal and formal mathematical reasoning and it will provide a more robust basis to prove the methods' positive effect on performance in the arithmetic development. Finally, we also deem that the effect of the ABN approach could be compared to other types of flexible or innovative approaches in the field of early mathematics teaching, in order to provide more support to the instructional potential of the ABN approach.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Bioethics Committee of University of Cádiz, available at http://www.uca.es/recursos/doc/Unidades/normativa/investigacion/122607032_151201413525.pdf. It was also carried out in accordance with the National Commission for Scientific and Technological Research (CONICYT, Chile) <http://www.conicyt.cl/fondecyt/category/estudios-y-documentos/bioetica/>. The protocol was approved by the Bioethics Committee of the three participant universities. All subjects gave written informed consent in accordance with the Declaration of Helsinki and Singapore Statement. http://www.conicyt.cl/fondecyt/files/2017/06/Singapore_Statement.pdf.

AUTHOR CONTRIBUTIONS

GC statistical analysis, literature review, and writing the final version of the manuscript. EA application of assessment tools in educational establishments and statistical analysis of results. CP theoretical discussion of instructional methods, fundamentals of the ABN methodology and contextualization of statistical results according the ABN focus. JN theoretical discussion of Early Mathematical competences, analysis of comparable results in literature, and review of the draft version of the manuscript. MA theoretical discussion of psychometric instruments applied. Application of assessment tools in educational establishments, responsible for the ethical consent of the participants.

FUNDING

This work has been partially funded by the projects FONDECYT 1160980, FONDECYT 11150201, Basal Financing Program FB0003 from the Associative Research Program at CONICYT, Chile, and PSI2015-63856-P of MINECO, Spain.

REFERENCES

- Adamuz-Povedano, N., and Bracho-López, R. (2014). Algoritmos flexibles para las operaciones básicas como modo de favorecer la inclusión social. *Rev. Int. Educ. Para Just. Soc.* 3, 37–53.
- Alsina, A. (2012). Más allá de los contenidos, los procesos matemáticos en educación infantil. *Edma 0-6: Educ. Matemát. Infancia* 1, 1–14.
- Aragón, E., Canto, M., Marchena, E., Navarro, J. I., and Aguilar, M. (2017a). Perfil cognitivo asociado al aprendizaje matemático con el método algoritmo abierto basado en números. *Rev. Psicod.* 22, 54–59. doi: 10.1016/S1136-1034(17)30044-8
- Aragón, E., Delgado, C., and Marchena, E. (2017b). Diferencias de aprendizaje matemático entre los métodos de enseñanza ABN y CBC. *Psychol. Soc. Educ.* 2017, 61–70. doi: 10.25115/psye.v9i1.462
- Araújo, A., Aragon, E., Aguilar, M., Navarro, J. I., and Ruiz, G. (2014). Un estudio exploratorio para la adaptación de la versión española revisada del Early Numeracy Test-R para evaluar el aprendizaje matemático temprano. [An exploratory study for the standardization of the Spanish version of «Early Numeracy Test-R» to mathematical learning assessment]. *Eur. J. Educ. Psychol.* 7, 83–93. doi: 10.1989/ejep.v7i2.181
- Barba-Uriach, D., and Calvo, C. (2012). Algoritmos antiguos de cálculo. *Cuadernos Pedagogía* 421, 62–65.
- Baroody, A. J., and Rosu, L. (2006). “Adaptive expertise with basic addition and subtraction combinations: the number sense view,” in *Developing Adaptive Expertise in Elementary School Arithmetic. Symposium on The annual meeting of the American Educational Research Association*, eds A. J. Baroody and T. Torbeyns, San Francisco, CA.
- Bartelet, D., Vaessen, A., Blomert, L., and Ansari, D. (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *J. Exp. Child Psychol.* 117, 12–28. doi: 10.1016/j.jecp.2013.08.010
- Bracho-López, R., Adamuz-Povedano, N., Jiménez-Fanjul, N., and Gallego-Espejo, M. C. (2014a). “Una experiencia de investigación-acción colaborativa para el desarrollo del sentido numérico en los primeros años de aprendizaje matemático,” in *Investigaciones en Pensamiento Numérico y Algebraico e Historia de las Matemáticas y Educación Matemática -2014*, eds J. L. En González et al. (Málaga: Departamento de Didáctica de las Matemáticas, de las Ciencias Sociales y de las Ciencias Experimentales y SEIEM).
- Bracho-López, R., Gallego-Espejo, M. D. C., Adamuz-Povedano, N., and Jiménez-Fanjul, N. (2014b). Impacto escolar de la metodología basada en algoritmos ABN en niños y niñas de primer ciclo de Educación Primaria. *Rev. Iberoamer. Educ. Matemát.* 39, 97–109.
- Bruner, J. S. (1966). *Toward a Theory of Instruction*. Cambridge: Harvard University Press.
- Cantos, S. (2016). *Entrevista a Jaime Martínez-Montero. La Voz de Cádiz*. Available at: <http://goo.gl/Tt5821>
- Cantrell, L., and Smith, L. B. (2013). Open questions and a proposal: a critical review of the evidence on infant numerical abilities. *Cognition* 128, 331–352. doi: 10.1016/j.cognition.2013.04.008
- Cargnelutti, E., Tomasetto, C., and Passolunghi, M. C. (2017). The interplay between affective and cognitive factors in shaping early proficiency in mathematics. *Trends Neurosci. Educ.* 8, 28–36. doi: 10.1016/j.tine.2017.10.002
- Cerdea, G., Pérez, C., Navarro, J. I., Aguilar, M., Casas, J., and Aragón, E. (2015). Explanatory model of emotional-cognitive variables in school mathematics performance: a longitudinal study in primary school. *Front. Psychol.* 6:1363. doi: 10.3389/fpsyg.2015.01363
- Chu, F. W., vanMarle, K., and Geary, D. C. (2015). Early numerical foundations of young children’s mathematical development. *J. Exp. Child Psychol.* 132, 205–212. doi: 10.1016/j.jecp.2015.01.006
- Cirino, P. T. (2011). The interrelationships of mathematical precursors in kindergarten. *J. Exp. Child Psychol.* 108, 713–733. doi: 10.1016/j.jecp.2010.11.004
- De Smedt, B., Noël, M.-P., Gilmore, C., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children’s mathematical skills? A review of evidence from brain and behavior. *Trends Neurosci. Educ.* 2, 48–55. doi: 10.1016/j.tine.2013.06.001
- De Smedt, B., Verschaffel, L., and Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *J. Exp. Child Psychol.* 103, 469–479. doi: 10.1016/j.jecp.2009.01.010
- Defever, E., Reynvoet, B., and Gebuis, T. (2013). Task- and age-dependent effects of visual stimulus properties on children’s explicit numerosity judgments. *J. Exp. Child Psychol.* 116, 216–233. doi: 10.1016/j.jecp.2013.04.006 doi: 10.1016/j.jecp.2013.04.006
- Fan, L., Zhu, Y., and Miao, Z. (2013). Textbook research in mathematics education: development status and directions. *ZDM* 45, 633–646. doi: 10.1007/s11858-013-0539-x
- Fawcett, A. J., and Nicolson, R. L. (2013). *DST- J: Test para la Detección de la Dislexia en Niños*. 3rd Ed. Madrid: Tea ediciones.
- Fazio, L. K., Bailey, D. H., Thompson, C. A., and Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *J. Exp. Child Psychol.* 123, 53–72. doi: 10.1016/j.jecp.2014.01.013
- Fuhs, M. W., and Mcneil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: contributions of inhibitory control. *Dev. Sci.* 16, 136–148. doi: 10.1111/desc.12013
- Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year long longitudinal study. *Dev. Psychol.* 47, 1539–1552. doi: 10.1037/a0025510
- Geary, D. C., Baily, D. H., Littlefield, A., Wood, P., Hoard, M. K., and Nugent, L. (2009). Firstgrade predictors of mathematical learning disability: a latent class trajectory analysis. *Cognit. Dev.* 24, 411–429. doi: 10.1016/j.cogdev.2009.10.001
- Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., et al. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS One* 8:e67374. doi: 10.1371/journal.pone.0067374.g002
- Ginsburg, H., and Baroody, A. J. (2007). *TEMA-3: Test de Competencia Matemática Básica*. Madrid, España: TEA ediciones.
- Ginsburg, H. P., Klein, A., and Starkey, P. (1998). “The development of children’s mathematical thinking: connecting research with practice,” in *Handbook of Child Psychology: Child Psychology in Practice*, Vol. 4, (Series Ed.), eds I. E. Sigel, K. A. Renninger, and W. Damon (New York, NY: Wiley), 401–476.
- Godino, J. D., and Font, V. (2003). *Razonamiento Algebraico Para Maestros*. Granada: Universidad de Granada.
- Hadar, L. L. (2017). Opportunities to learn: mathematics textbooks and students’ achievements. *Stud. Educ. Eval.* 55, 153–166. doi: 10.1016/j.stueduc.2017.10.002
- Harvey, H. A., and Miller, G. E. (2017). Executive function skills, early mathematics, and vocabulary in Head Start preschool children. *Early Educ. Dev.* 28, 290–307. doi: 10.1080/10409289.2016.1218728
- Hernández, R., Fernández, C., and Baptista, P. (2014). *Metodología de la Investigación*, 6th Edn. México D.F: McGraw-Hill.
- Humberstone, J., and Reeve, R. A. (2018). The conceptual overlap between arithmetic and algebraic referential mapping. *Learn. Instruct.* 54, 138–146. doi: 10.1016/j.learninstruc.2017.09.001
- Jordan, N. C., Kaplan, D., Locuniak, M. N., and Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learn. Disabil. Res. Pract.* 22, 36–46. doi: 10.1111/j.1540-5826.2007.00229.x
- Jordan, N. C., Kaplan, D., Ramineni, C., and Locuniak, M. N. (2008). Development of number combination skill in the early school years: when do fingers help? *Dev. Sci.* 11, 662–668. doi: 10.1111/j.1467-7687.2008.00715.x
- Jordan, N. C., Kaplan, D., Ramineni, C., and Locuniak, M. N. (2009). Early math matters: kindergarten number competence and later mathematics outcomes. *Dev. Psychol.* 45, 850–867. doi: 10.1037/a0014939
- Laski, E. V., and Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Dev.* 78, 1723–1743. doi: 10.1111/j.1467-8624.2007.01087.x
- Levitin, A. (2015). *Introduction to the Design & Analysis of Algorithms*. Boston, CA: Addison-Wesley.
- Lo, M.-T., Chen, S.-K., and Lin, S. (2017). Groups holding multiple achievement goals in the math classroom: profile stability and cognitive and affective outcomes. *Learn. Ind. Diff.* 57, 65–76. doi: 10.1016/j.lindif.2017.06.001

- Lyons, I., Price, G., Vaessen, A., Blomert, L., and Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Dev. Sci.* 17, 714–726. doi: 10.1111/desc.12152
- Mancebon-Torrubia, M. J., and Perez Ximenez-de-Embun, D. (2014). Equality of school choice: a study applied to the Spanish region of Aragón. *Educ. Econ.* 22, 90–111. doi: 10.1080/09645292.2010.545197
- Martínez-Montero, J. (2010). *Enseñar Matemáticas a Estudiantes Con Necesidades Educativas Especiales*, 2nd Edn. Barcelona: CISS-Praxis.
- Martínez-Montero, J. (2011). El método de cálculo abierto basado en números (ABN) como alternativa de futuro respecto a los métodos tradicionales cerrados basados en cifras (CBC). *Bordón* 63, 95–110.
- Martínez-Montero, J., and Sánchez, C. (2011). *Desarrollo y Mejora de la Inteligencia Matemática en Educación Infantil*. Madrid: Wolters Kluwer S.A.
- Martínez-Montero, J., and Sánchez, C. (2013). *Resolución de Problemas y Método ABN*. Madrid: Wolters Kluwer Educación.
- Moore, A. M., vanMarle, K., and Geary, D. C. (2016). Kindergartners' fluent processing of symbolic numerical magnitude is predicted by their cardinal knowledge and implicit understanding of arithmetic 2 years earlier. *J. Exp. Child Psychol.* 150, 31–47. doi: 10.1016/j.jecp.2016.05.003
- Mou, Y., and vanMarle, K. (2014). Two core systems of numerical representation in infants. *Dev. Rev.* 34, 1–25. doi: 10.1016/j.dr.2013.11.001
- National Governors Association Center for Best Practices and Council of Chief State School Officers (2010). *Common Core State Standards for Mathematics*. Washington, DC: Authors.
- Navarro, J. I., Aguilar, M., Marchena, E., Ruiz, G., Menacho, I., and Van Luit, H. (2012). Longitudinal study of low and high achievers in early mathematics. *Br. J. Educ. Psychol.* 82, 28–41. doi: 10.1111/j.2044-8279.2011.02043.x
- Novo, M. L., Alsina, Á., Marbán, J. M., and Berciano, A. (2017). Connective intelligence for childhood mathematics education. *Comunicar* 25:29. doi: 10.3916/C52-2017-03
- Núñez, M. C., and Lozano, I. (2009). Evaluación del progreso en competencia matemática básica. Estudio de casos a través del TEMA-3: alumnos con y sin discapacidad psíquica. *Indiv. Bol. Estud. Investig. Monogr.* 12, 139–160.
- OECD (2015). *What Makes Schools Successful? Resources, Policies and Practices*. Available at: <https://www.oecd.org/pisa/keyfindings/Vol4Ch2.pdf>
- Peng, P., Namkung, J., Barnes, M., and Sun, C. (2016). A meta-analysis of mathematics and working memory: moderating effects of working memory domain, type of mathematics skill, and sample characteristics. *J. Educ. Psychol.* 108, 455–473. doi: 10.1037/edu0000079
- Robinson, K. M., and Dubé, A. K. (2012). Children's use of arithmetic shortcuts: the role of attitudes in strategy choice. *Child Dev. Res.* 2010:10. doi: 10.1155/2012/459385
- Robinson, K. M., Price, J., and Demyen, B. (2018). Understanding arithmetic concepts: does operation matter? *J. Exp. Child Psychol.* 166, 421–436. doi: 10.1016/j.jecp.2017.09.003
- Ryoo, J. H., Molfase, V., Brown, T., Karp, K., Welch, G., and Bovaird, J. (2015). Examining factor structures on the test of early mathematics ability-3: a longitudinal approach. *Learn. Ind. Diff.* 41, 21–29. doi: 10.1016/j.lindif.2015.06.003
- Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi: 10.1111/j.1467-8624.2004.00684.x
- Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–243. doi: 10.1111/1467-9280.02438
- Starr, A., DeWind, N., and Brannon, E. (2017). The contributions of numerical acuity and non-numerical stimulus features to the development of the number sense and symbolic math achievement. *Cognition* 168, 222–233. doi: 10.1016/j.cognition.2017.07.004
- Szkudlarek, E., and Brannon, E. (2017). Does the approximate number system serve as a foundation for symbolic mathematics? *Lang. Learn. Dev.* 13, 171–190. doi: 10.1080/15475441.2016.1263573
- Szucs, D., Nobes, A., Devine, A., Gabriel, F. C., and Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Front. Psychol.* 4:44. doi: 10.3389/fpsyg.2013.00444
- Torbeys, J., Verschaffel, L., and Ghesquiere, P. (2005). Simple addition strategies in a first grade class with multiple strategy instruction. *Cognit. Instruct.* 23, 1–121. doi: 10.1207/s1532690xcic2301_1
- Vogel, S., Remark, A., and Ansari, D. (2015). Differential processing of symbolic numerical magnitude and order in first grade children. *J. Exp. Child Psychol.* 129, 26–39. doi: 10.1016/j.jecp.2014.07.010
- Wang, L., Sun, Y., and Zhou, X. (2016). Relation between approximate number system acuity and mathematical achievement: the influence of fluency. *Front. Psychol.* 7:1966. doi: 10.3389/fpsyg.2016.01966
- Watts, T. W., Duncan, G. J., Siegler, R. S., and Davis-Kean, P. E. (2014). What's past is prologue: relations between early mathematics knowledge and high school achievement. *Educ. Res.* 43, 352–360. doi: 10.3102/0013189X14553660
- Wechsler, D. (2009). *Test de Inteligencia Para Niños Preescolares–WPPSI III*. Madrid: TEA Ediciones.

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

Copyright © 2018 Cerdeá, Aragón, Pérez, Navarro and Aguilar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Associative Cognitive Factors of Math Problems in Students Diagnosed With Developmental Dyscalculia

Johannes Erik Harold Van Luit* and Sylke Wilhelmina Maria Toll

Department of Pedagogics and Education, Faculty of Social Sciences, Utrecht University, Utrecht, Netherlands

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

Robert Reeve,
The University of Melbourne, Australia
Sabine Heim,
Rutgers University, The State
University of New Jersey,
United States

*Correspondence:

Johannes Erik Harold Van Luit
j.e.h.vanluit@uu.nl

Specialty section:

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

Received: 24 January 2018

Accepted: 18 September 2018

Published: 09 October 2018

Citation:

Van Luit JEH and Toll SWM (2018)
Associative Cognitive Factors of Math
Problems in Students Diagnosed With
Developmental Dyscalculia.
Front. Psychol. 9:1907.
doi: 10.3389/fpsyg.2018.01907

The Dutch protocol, 'Dyscalculia: Diagnostics for Behavioral Professionals' (DDBP protocol; Van Luit et al., 2014), describes how behavioral experts can examine whether a student has developmental dyscalculia (DD), based on three criteria: severity, discrepancy, and resistance. In addition to distinguishing the criteria necessary for diagnosis, the protocol provides guidance on formulating hypotheses by describing and operationalising four possible associative cognitive factors of math problems: *planning skills*, *naming speed*, *short-term and/or working memory*, and *attention*. The current exploratory and descriptive research aims to describe the frequency of these four primary associative cognitive factors in students with DD from the Netherlands. Descriptive data from 84 students aged 8–18 years showed that deficits in *naming speed* (in particular, in naming numbers) were the most frequent explanation of math problems in children with DD, followed by deficits in *short-term/working memory* and *planning skills*. Deficits in *attention* were the least frequent. The findings are explained in light of current literature, and suggestions for follow-up research are presented.

Keywords: dyscalculia, planning, naming speed, memory, attention, diagnosis, protocol

INTRODUCTION

Many students in primary and secondary education experience problems with mathematics (Geary, 2004). Math problems can have major consequences for their further educational career and for their ability to live independently in society (Every Child a Chance Trust, 2009). Math problems that are extensive and persistent in nature may indicate developmental dyscalculia (DD). Although there is inconsistent use of terminology in the literature, researchers agree that DD refers to the existence of a severe disability in learning mathematics. Ruijsenaars et al. (2016, p. 28) defined DD as a disorder characterized by persistent problems with learning and fluency and/or accurate recall and/or application of mathematical knowledge (facts and understanding). The prevalence of DD is estimated to be between 2 and 3% in students in the Netherlands (Ruijsenaars et al., 2016). Percentages are higher in international research (3–8%), depending on how researchers define such mathematical disorders (Desoete et al., 2004; Dowker, 2005; Shalev et al., 2005). The disability can be highly selective, affecting learners with normal intelligence (e.g., Landerl et al., 2004), although it also co-occurs with other developmental disorders, including reading disorders (Ackerman and Dykman, 1995; Light and DeFries, 1995; Gross-Tsur et al., 1996) and attention deficit hyperactivity disorder (ADHD; Monuteaux et al., 2005).

Within the revised fourth edition of the Diagnostic and statistical manual of mental disorders (DSM-IV-TR; American Psychiatric Association, 2000) the now-obsolete diagnostic criteria for Mathematics Disorder (code: 315.1) were: (A) Mathematical ability, as measured by individually administered standardized tests, is substantially below that expected given the person's chronological age, measured intelligence, and age-appropriate education; (B) The disturbance in Criterion A significantly interferes with academic achievement or activities of daily living that require mathematical ability; and (C) If a sensory deficit is present, the difficulties in mathematical ability are in excess of those usually associated with it. The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013) takes a different approach to learning disorders than previous editions of the manual by broadening the category, in order to increase diagnostic accuracy and effectively target care. Specific learning disorder is now a single, overall diagnosis, incorporating various deficits that impact academic achievement. The criteria describe shortcomings in general academic skills, providing detailed specifiers for the areas of reading, mathematics, and written expression. Diagnosis of this disorder requires persistent difficulties in reading, writing, arithmetic, or mathematical reasoning skills during the formal years of schooling. Symptoms may include inaccurate or slow and effortful reading, poor written expression that lacks clarity, difficulties remembering number facts, or inaccurate mathematical reasoning. Current academic skills must be well below the average range of scores in culturally and linguistically appropriate tests of reading, writing, or mathematics. The individual's difficulties must not be better explained by developmental, neurological, sensory (vision or hearing), or motor disorders and must significantly interfere with academic achievement, occupational performance, or activities of daily living.

Despite the changes from DSM-IV-TR to DSM-5, it remains necessary to perform extensive diagnostic testing to establish whether DD is present. The Dutch 'Protocol dyscalculia: Diagnostics for behavioral professionals' (DDBP protocol; Van Luit et al., 2014) describes how behavioral experts can examine whether a student, from 8 years of age and older, has DD. The DDBP protocol contains guidelines and suggestions about the variables that can be investigated, and the methods used, during a diagnostically examination of DD. Due to its structured and comprehensive nature, the DDBP protocol has now been systematically implemented in many social care settings in education in the Netherlands and Flanders (Belgium). The DDBP protocol deals with the criteria that must be met in order to diagnose DD (Van Luit, 2012; Van Luit et al., 2014), namely:

(1) *The criterion of severity*: there is a significant delay in automated math skills as compared to peers and/or fellow children and a significant delay in mastering the substantive math skills of the various domains. At the end of primary school, for example in sixth grade, there must be a delay of at least 2 years on a standard (national) math test. For such a test this would mean that a student at the end of sixth grade would fail a test designed for children at the end of fourth grade. In earlier grades, for example halfway through fifth grade, this would mean that the

student would fail the test designed for students at the end of third grade. At the beginning of fourth grade, a student would fail the test designed for children at the end of second grade. Dyscalculia is rarely diagnosed before the end of third grade.

(2) *The criterion of discrepancy*: there is a significant delay in mathematics with respect to what can be expected of the individual, based on their individual development. In determining dyscalculia, the presence of an average intelligence is not typically required. The cognitive level is mostly assessed by an intelligence test. Children with dyscalculia can have an under- or above-average intelligence level. It is not possible to determine dyscalculia when the student has an intelligence score of 70 or below, because in that case the low mathematical skills are expected relative to the child's personal abilities. When the total IQ score is between 71 and 85, diagnosing dyscalculia must be done with caution. Mathematics requires a complex skill set that relies on higher cognitive functions. Therefore, it is not realistic to expect that children with an IQ at this level will develop and achieve the same math abilities as their peers with an average IQ score. For these children the lag in mathematical skills needs to be larger (at the end of grade six, at least three years) than the lag of mathematical skills of a person with an average intelligence score (at the end of grade six, at least 2 years).

(3) *The criterion of didactic resistance*: there is a persistent mathematical problem, which is resistant to specialized help. To determine the persistence of the deficit, the structural and specialized help a student had received in mathematics is investigated. Receiving most attention here are past reports of offered help. According to the model of 'response-to-instruction,' didactic resistance can only be determined with full certainty when the conditions for all three criteria have been met (Fuchs and Vaughn, 2012). Thus dyscalculia cannot be diagnosed if the third criterion has not been complied with, a condition that also applies to children in secondary school.

Since recent research has increasingly recognized the heterogeneity of DD by differentiating among underlying cognitive deficits (Murphy et al., 2007; Rubinsten and Henik, 2009; Geary, 2011; Kaufmann et al., 2013; Skagerlund and Träff, 2016), identification of DD does not on its own provide sufficient information about the educational needs of an individual student with math problems. The DDBP protocol therefore provides, in addition to the above three criteria, guidance on performing diagnostic research by describing and operationalising four possible associative, or primary, factors related to a student's math problems: *planning skills*, *naming speed*, *short-term and working memory*, and *attention*. (Number sense is also mentioned within the DDBP protocol, but was not taken into account in this research due to issues relating to the time frame.) The five factors (including number sense) are in line with international research on the underlying neurocognitive correlational and causal factors in mathematical difficulties (Träff et al., 2017). Where the diagnosis of DD provides some information about the presence of the problem, identification of these additional factors enables a more complete and integrated picture of an individual student's educational needs – and thus a sounder basis for appropriate interventions. This might include compensation, remediation

and/or dispensation, depending on discovered associative factor(s).

A distinction is made here between primary associative and secondary associative factors. Secondary factors mentioned in the protocol, for example, are work attitude and motivation, self-concept, anxiety, reading problems and delayed or disturbed social-emotional development (e.g., Carey et al., 2016; Sorvo et al., 2017). As mentioned, the DDBP protocol names five primary associative factors; however, these possible factors are certainly not exhaustive. The first primary factor is *planning skills*. Planning processes are required during math tasks for choosing and applying strategies, monitoring calculation(s), applying mathematical knowledge, and checking the answer (Das and Naglieri, 1997). Deficits in planning skills therefore seem to explain an important part of why students with DD have difficulty performing mathematical procedures. Students with DD have been found to have deficits in planning, as compared to students without DD (Kroesbergen et al., 2003). The second primary factor is *naming speed*. Naming speed is the speed of access to (specific) information in long-term memory. A deficit in naming speed can mean that more time and effort is needed during math tasks to make relevant information readily available for solving a task. In students with DD, there is evidence of a deficit in the naming speed of numbers or, alternatively, general deficits in naming speed (D'Amico and Passolunghi, 2009; Mazzocco and Grimm, 2013; Koponen et al., 2017). The third primary factor is *short term and/or working memory*. During math tasks, a large amount of information must be retained and processed. This requires the application of both short-term and working memory. Various studies have found that difficulties in storing, editing, and reproducing auditory information in verbal memory (Berg, 2008) as well as visuospatial information in visual memory (D'Amico and Guarnera, 2005) can underlie deficits in DD (Raghubar et al., 2010; Toll et al., 2011). The fourth primary factor is *attention*. Being able to focus and maintain attention ensures that math tasks are accurately represented during problem solving and, further, that math facts are readily and accurately recalled from memory (Passolunghi and Cornoldi, 2000). Also, by maintaining attention, a student can focus on math problems for longer periods of time (Roeyers and Baeyens, 2016). In the case of students with DD, attentional skills are often weaker (Kroesbergen et al., 2003); these students also can have difficulty suppressing responses (i.e., lower inhibitory control; D'Amico and Passolunghi, 2009; Ashkenazi and Henik, 2010; Navarro et al., 2011). The fifth primary factor is *number sense*. Recent research (De Smedt et al., 2009; Fuchs et al., 2010; LeFevre et al., 2010; Kolkman et al., 2013; Schneider et al., 2017) shows that number sense, the ability to process, understand and estimate numerical quantities (Dehaene, 1992), is a predictive factor in the development of math skills. Deficits in number sense appear to be a possible explanation for serious math problems (Mussolin et al., 2010; Piazza et al., 2010; Mazzocco et al., 2011). Research in developmental neuroscience (e.g., Olsson et al., 2016) has even identified neural markers of impairments in numerosity processing in DD (for a review, see Butterworth et al., 2011).

Although recent research has repeatedly linked the five factors to the presence of math problems, few studies (e.g.,

Navarro et al., 2011) have tested all the factors together. The literature is also missing clinically oriented research involving an adequate number of students diagnosed with persistent mathematical learning disabilities (i.e., DD). The limited amount of research into DD is largely due to issues of feasibility and generalisability. However, research into a target clinical group, whether descriptive or not, can provide valuable information about the presence of the factors in students with DD. Of particular benefit can be information about the frequency with which those students fail to perform, in comparison with their peers.

The current research aimed to describe the frequency of four of the five primary associative factors in students with DD. Otherwise put, it examined in how many children the four factors¹ could be identified as underlying mechanisms of their math issues. The central research question was, 'What is the frequency of deficits in planning skills, naming speed, short-term and/or working memory, and attention in children with DD?' The research objective was twofold. Firstly, the study aimed to examine, per factor, the percentage of students showing deficits in that particular domain. Secondly, it sought to investigate the multifactorial distribution of deficits in these primary factors, as may contribute to DD, i.e., how many students with DD have deficits on one or more of the primary associative factors? Although insight into a limited group of children does not provide information about the strength of the relationship between an associative factor and the presence of math problems per se, it lends support for the outlined diagnostic framework in the DDBP protocol. This may be especially so because of the clinical sample itself and the extensive diagnostic evaluations performed prior to diagnosing DD – and thus prior to inclusion in this research. Empirical evidence regarding these underlying factors can provide additional insight into the nature of such mathematical deficits. As such, it can contribute to accountability of procedures within diagnostic care. This can help clinicians and teachers alike to identify targets of intervention and, as well, enable students with DD to overcome their deficiencies in the field of mathematics.

MATERIALS AND METHODS

Participants

The participants included in this study were 84 students from the Netherlands (8–18 years of age) were visiting a university institute for learning difficulties because of their problems with mathematics and were not diagnosed with DD before. Only those who were diagnosed with DD were included in this study. Diagnostic examination into math difficulties was conducted during the period 2009–2015, and was based on diagnostic research using the three criteria of the DDBP protocol. A consent

¹The fifth factor, number sense, was not included in this research because of the time frame of the study (in the DDBP protocol, number sense has only been considered a primary factor since 2014; Van Luit et al., 2014). This meant that, first, number sense skills had been measured in too few children and, second, too much variation between test materials made comparison of the available data on number sense impossible.

statement for participation in scientific research was signed by parent(s) or those with parental authority for all participants. **Table 1** describes the 84 students included in the current research. More than three quarters of the students ($n = 64$, 76.2%) were girls. This skewed distribution of sex was remarkable because previous research (Devine et al., 2013) did not reveal such gender differences. The mean intelligence score of the participants was 91.28 ($SD = 11.30$). For almost all students ($n = 81$), the Speed Number Facts Test (SNFT; De Vos, 1992, 2010) had been administered as one component of the diagnostic procedure. The SNFT was used to measure the amount of memorized mathematical facts. The raw average total score on the SNFT was 73.00 ($SD = 30.83$, $n = 58$) on the 2010 version and 67.00 ($SD = 20.63$, $n = 20$) on the 1992 version. In the case of three students, due to their young age and in accordance with the manual, only the addition and subtraction parts were administered. Their average raw score on the 2010 version was 27.50 ($SD = 0.71$, $n = 2$), and on the 1992 version was 19.00 ($n = 1$).

Procedure

All students had come for diagnostic assessment at the same university institute for learning difficulties. The child's scores, as obtained from the diagnostic examination, were anonymously processed in an SPSS database. In order to diagnose DD, the following procedure was followed: (a) collection of information on (academic) performance from parents and school records (answers on standardized questionnaires and data from national mathematics tests); and (b) individual diagnostic examination (administered in two to three blocks of time, each lasting approximately 5 h). All diagnostic testing took place individually in a quiet space and was performed by a clinician with at least a master's degree in psychology under supervision of a psychologist with a doctoral degree.

Measures

The performance of the students on each primary factor was measured with one specific instrument (as part of the detailed procedure described in DDBP). A description of each instrument is provided below. Due to diagnostic considerations (age, time, the size of the test battery, etc.), not every instrument was used with every student. For each research measure, we indicated how many student scores were available. Seventy students (83.3%) were administered all measures. The instrument descriptions below include information on standardization (mean and standard deviation), and a cut-off score (mean minus one standard deviation) that indicates deficits in the specific area.

Planning

The Planning scale of the Cognitive Assessment System (CAS; individual test for children aged 5–17 years; Das and Naglieri, 1997) was used to measure *planning skills*. This scale consists of two (short version) or three (full version) subtests. In both cases a standardized score is derived ($M = 100$, $SD = 15$). The subtests are “Matching Numbers,” “Planned Codes,” and “Planned Connections” (Das and Naglieri, 1997). The “Matching Numbers” subtest consists of four pages, each with eight rows of six numbers. The numbers increase in size from one to seven digits. The student must underline the two corresponding numbers in each row. The “Planned Codes” subtest consists of two parts. A legend at the top of the page shows which codes belong to the letters A through D. The page contains 56 letters without codes arranged in different combinations. The student must fill in the correct code below each letter. The “Planned Connections” subtest consists of items that increase in difficulty. Each item consists of a page where numbers or letters, distributed randomly across the page, must be connected by the student in the correct order. The subtest scores are determined by both accuracy and speed (Das and Naglieri, 1997). A score below 85 indicates deficits in planning. The Planning scale was administered to 81 students (96.4%). The CAS has been found to provide a valid picture of information processing (Kroesbergen et al., 2002; Van Luit et al., 2005); the average reliability coefficient for the Planning scale is 0.88 (Das and Naglieri, 1997; Naglieri, 1999).

Naming Speed

Four cards of the Rapid Naming & Reading Test (RN&RT; Van den Bos and Lutje Spelberg, 2007) were used to measure the *naming speed* of colors, digits, pictures and letters. This task provides an indication of how quickly a student can extract verbal information about visual characters from memory. With each card, the student must identify one kind of visual character as quickly as possible. The time (in seconds) a student takes to name the characters on a single card is counted as the raw score; this is then converted to a standard score ($M = 10$, $SD = 3$). A standard score below seven indicates deficits in naming speed. The RN&RT was administered to 80 students (95.2%). Research has shown that the RN&RT is sufficiently reliable and valid (Van den Bos and Lutje Spelberg, 2007).

Short-Term and Working Memory

The computerized Automated Working Memory Assessment (AWMA; Alloway, 2007) was used to measure the capacity of

TABLE 1 | Participant data for total group by education level.

	<i>n</i> (%)	Sex		Age in months			
		Boys (%)	Girls (%)	<i>M</i>	<i>SD</i>	Min	Max
Primary education	56 (66.7)	14 (25.0)	42 (75.0)	126.21	12.90	96	149
Other education*	28 (33.3)	6 (21.4)	22 (78.6)	175.00	18.73	145	212
Total	84 (100.0)	20 (23.8)	64 (76.2)	142.48	27.56	96	212

*Secondary or vocational education (younger than 18 years).

short-term and working memory. Scores on the AWMA are fairly stable during the primary school period and show good convergence with WISC-IV memory tasks (Alloway et al., 2008). Each subtest out of six starts with a practice session and consists of blocks containing six trials. The first block of each trial consists of one stimulus. For each subsequent block, the trial increases by one stimulus. After three errors in one block, the task is terminated. After four correct answers in one block, the student can proceed to the next block and receives a maximum six points. In other cases, the score is the number of correct items per block. The raw score is the sum of all points scored. This raw score is converted to a standard score ($M = 100$, $SD = 15$) per subtest. The AWMA differentiates between four components of short-term and working memory. Verbal short-term memory is measured with the subtests “Digit recall,” “Word recall” and/or “Non-word recall.” In these three subtests the student hears a series of verbally presented numbers, words and/or nonsense words, and must then recall this series correctly. If more than one of the subtests was administered, an average for the verbal short-term memory was calculated based on information in the manual. Verbal working memory is measured with the subtest “Listening recall.” The student hears a series of spoken sentences, and at the end of each series must: (a) indicate whether the sentence is true or false, and (b) recall the last word of each sentence in sequence. (The true/false judgment is not included in the scores.) Visuospatial short-term memory is measured with the subtest “Dot matrix.” The student is shown the position of red dots in a matrix of 4×4 boxes for two seconds. The position of these dots must be identified in the correct order after the dots have disappeared. Visuospatial working memory is measured with the subtest “Odd one out.” The student views three shapes in boxes next to each other and identifies the shape different from the others. At the end of each trial, the student must identify, in the correct order, the location of each shape that was the odd-one-out. A standard score lower than 85 indicates deficits in memory. Deficits may occur on one specific memory component or multiple components at a time. The test–retest reliability for the subtests is, respectively, 0.84, 0.76, 0.64, 0.81, 0.83, and 0.81 (Alloway et al., 2006).

Attention

The Attention scale of the Cognitive Assessment System (CAS; individual test for children aged 5–17 years; Das and Naglieri, 1997) was used to measure *attention*. The scale consists of two (short version) or three (full version) subtests. In both cases a score was calculated ($M = 100$, $SD = 15$). The subtests are “Expressive Attention,” “Number Detection,” and “Receptive Attention” (Das and Naglieri, 1997). The “Expressive Attention” subtest consists of a page with words like “blue” and “red” printed in different colors. The student must name the color in which the words are printed; the dominant response, the word which is read, must be suppressed. Two exercises are taken in advance to determine whether the student is sufficiently capable of naming words and colors. The subtest score is determined based on speed and accuracy on the final task. The “Number Detection” subtest consists of two pages with numbers. These numbers are printed in different fonts. On each page the student must underline the

numbers that look the same as those at the top of the page (e.g., 1, 2, 3 printed in open font). This requires selectively focusing attention on specifically printed numerical symbols. The “Receptive Attention” subtest consists of pictures or letters in pairs. The student must underline when the two pictures/letters are the same or have a similar characteristic. These subtest scores are also determined by accuracy and speed (Das and Naglieri, 1997). A scale score below 85 indicates deficits in attention. The Attention scale was administered to 79 students (94.0%). The CAS provides a valid picture of information processing (Kroesbergen et al., 2002; Van Luit et al., 2005), and the average reliability coefficient for the Attention scale is 0.88 (Das and Naglieri, 1997; Naglieri, 1999).

RESULTS

Table 2 shows descriptive statistics for each factor and component. On nearly all measures the average participant scores were less than the mean scale or standard score (100 for planning, memory and attention, 10 for naming speed), though not lower than the criterion score indicating deficits in these factors (<85 for planning, memory and attention, <7 for naming speed). In **Table 3** correlations between all factors and components are presented. This table shows significant correlations between all factors (e.g., at least each factor correlated significantly with at least one other factor). A strong association ($r > 0.50$) was found between *planning skills* and *attention*, and between components within *naming speed*, i.e., colors-pictures, numbers-pictures, and letters-pictures). Moderate to strong associations ($0.3 < r < 0.50$) were found between *planning skills* and *naming speed* (i.e., colors); *naming speed* (i.e., colors) and *short-term/working memory* (i.e., visual STM); and *naming speed* (i.e., colors) and *attention*.

Table 4 gives an overview of the number of students with deficits by factor and component. Deficits in *naming speed* were found in 54 students (64.3%), deficits in *short-term/working memory* in 41 students (49.4%), deficits in *planning skills* in 37 students (45.7%) and deficits in *attention* in 10 students (12.7%). Within the *naming speed* factor, deficits in naming numbers were the most common ($n = 37$, 46.3%) and deficits in naming colors were the least common ($n = 26$, 32.5%). Within the *short-term/working memory* factor, deficits in visual short-term memory were the most common ($n = 21$, 25.0%) and deficits in verbal working memory were the least common ($n = 4$, 4.9%). The number of students with deficits was associated with specific factors; there were significantly more students with deficits in *planning skills*, *naming speed* and *short-term/working memory* than in *attention* [$\chi^2(9) = 69.63$, $p < 0.01$].

Table 5 gives an overview of the distribution of deficits in the factors across students. To enhance clarity, in this table the components are integrated into information about the given factor. In other words, a deficit in one of the elements (subtests) comprising *naming speed* or *planning skills* has been considered as a deficit in that factor as a whole (instead of separate components within that factor). The first column shows the number of deficits (zero up to four) that could be present in

students. The second column presents the number of students who experienced each deficit. In the remaining columns the contribution of the deficits over the four factors is shown. For

example, **Table 5** shows that only one primary factor was found in 26 students (31.0%). Within these 26 pupils, 61.5% had deficits in *naming speed*, 23.1% had deficits in *short term and/or working*

TABLE 2 | Descriptive statistics for each factor and component.

Factor	Component	<i>n</i>	<i>M</i>	<i>SD</i>	Min	Max
Planning						
	Planning	81	86.47	12.77	57.00	127.00
Naming speed						
	Colors	80	8.58	3.39	1.00	16.00
	Numbers	80	8.36	3.36	2.00	16.00
	Pictures	80	8.68	3.30	1.00	16.00
	Letters	80	9.16	3.28	1.00	17.00
Memory						
	Verbal STM	81	96.07	13.37	68.00	132.00
	Verbal WM	82	109.72	14.31	74.00	138.00
	Visuospatial STM	80	98.75	18.53	62.00	141.00
	Visuospatial WM	79	96.92	14.54	65.00	139.00
Attention						
	Attention	79	95.81	11.65	63.00	133.00

STM, short-term memory; WM, working memory.

TABLE 3 | Correlations between all factors and component.

Factor (– component)	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
(1) Planning	–	0.41**	0.21	0.30**	0.16	0.28*	0.11	0.28*	0.20	0.60**
(2) Naming speed – colors		–	0.49**	0.68**	0.50**	0.18	0.21	0.37**	0.34**	0.46**
(3) Naming speed – numbers			–	0.55**	0.66**	0.01	–0.04	0.02	0.23*	0.26*
(4) Naming speed – pictures				–	0.60**	0.30**	0.23*	0.33**	0.41**	0.36*
(5) Naming speed – letters					–	0.14	0.04	0.16	0.22	0.25*
(6) Memory – Verbal STM						–	0.32**	0.09	0.32**	0.27*
(7) Memory – Verbal WM							–	0.21	0.40**	0.13
(8) Memory – Visual STM								–	0.43**	0.29*
(9) Memory – Visual WM									–	0.33**
(10) Attention										–

* $p < 0.05$, ** $p < 0.01$, STM, short-term memory; WM, working memory.

TABLE 4 | Numbers and percentages of students with deficits for each factor and component.

Factor	Component	Deficit		No deficit	
		<i>n</i>	%	<i>n</i>	%
Planning		81	37	44	54.3
Naming speed		80	54	26	32.5
	Colors	80	26	54	67.5
	Numbers	80	37	43	53.8
	Pictures	80	31	49	61.3
	Letters	80	27	53	66.3
Memory		83	41	42	50.6
	Verbal STM	81	15	66	81.5
	Verbal WM	82	4	78	95.1
	Visuospatial STM	80	21	59	73.8
	Visuospatial WM	79	16	63	79.7
Attention		79	10	69	87.3

STM, short-term memory; WM, working memory.

TABLE 5 | Overview of the distribution of deficits in primary factors.

Number of deficits	Students with this number of deficits		Deficit in factor							
			Planning		Naming speed		Memory		Attention	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
0	13	15.5	0	0.0	0	0.0	0	0.0	0	0.0
1	26	31.0	4	15.4	16	61.5	6	23.1	0	0.0
2	25	29.8	14	56.0	19	76.0	17	68.0	0	0.0
3	14	16.7	13	92.9	13	92.9	12	85.7	4	28.6
4	6	7.1	6	100.0	6	100.0	6	100.0	6	100.0

memory and 15.4% had deficits in *planning skills*. **Table 5** shows three findings. Firstly, in 13 of 84 pupils, no primary factor of DD was found. Secondly, the table shows that *naming speed* was the most common unique factor of mathematical problems, with 61.5% of students with a deficit in at least one component within naming speed. Thirdly, *attention* did not occur as a unique factor, but only in conjunction with at least two other primary factors of mathematical problems.

DISCUSSION

The purpose of the current study was to describe the frequency of deficits on four primary associative factors for students with a diagnosis of DD: *planning*, *naming speed*, *short-term/working memory*, and *attention*. In 84 students aged 8–18 years with a diagnosis of DD (according to the DDBP protocol), the presence of deficits in these four factors was explored. Descriptive information showed that no primary factor of DD was found in 15.5% of the students. According to the DDBP protocol, establishing DD with certainty is difficult when the underlying factors of the mathematical problems remain unclear (Van Luit et al., 2014). The protocol indicates that, if no primary cause is found, a combination of secondary associative factors may also lead to the diagnosis. For those student participants with a diagnosis of DD but no underlying cognitive deficit, it may be that: (a) number sense served as an important factor when this factor is developed weak; (b) there were sufficient secondary factors that supported the diagnosis; (c) the student had above-average intelligence, meaning the criteria of deficits in *planning*, *naming speed*, *short-term/working memory* and/or *attention* were compensated for by other cognitive strengths; and/or (d) other primary associative factors, as yet not identified, played a role. Indeed, research has not yet been sufficiently conclusive to confirm that there are only five primary associative factors underlying DD (Van Luit et al., 2014; Träff et al., 2017).

The first research goal was to determine the percentage of students with deficits in the primary factors, indicating deficits in specific skills. The results show that deficits in naming speed (especially in naming numbers) were diagnosed most frequently, followed by deficits in short-term and working memory and in the field of planning. Deficits in attention were diagnosed least frequently. This could be explained in part by the fact that, in children with (probable) AD(H)D, research into DD typically

does not take place before the symptoms of this disorder have been reduced due to therapy and/or medication.

The second research goal was to investigate the multifactorial distribution of deficits across these four primary factors in students with DD. Deficits on one or two associative factors were found for most students. In 15.5% of the students, no primary underlying factor for DD was found. In 31% of the students one primary underlying factor was found and in the remaining 53.6% of the children, deficits on three or four factors were found. A breakdown of the four factors provided two interesting insights into the presence of the four primary associative factors in students with DD. Firstly, naming speed was the most common unique factor of math problems. In more than half the cases (61.5%), students with DD had difficulty readily finding relevant information for solving a task. This finding is consistent with the results from studies of D'Amico and Passolunghi (2009), and Mazzocco and Grimm (2013). Secondly, in the current study, attentional deficits never appeared as a unique primary factor in pupils with DD, a finding also shown in previous research (e.g., Roeyers and Baeyens, 2016). Deficits in attention were only found when at least two other primary factors were present, and this was the case only for a small portion of the sample (11.9%). This means that focusing and sustaining attention can play a role in DD, but these abilities are not at the forefront for this particular clinical target group. As noted earlier, this finding is possibly due to the deferral of children diagnosed with (probable) AD(H)D. Deficits in *naming speed*, *short-term/working memory* and *planning* were found to be the factors occurring most frequently in the student participants.

Although this research was exploratory and descriptive, by differentiating among cognitive deficits as may underly DD, our findings nevertheless support the line of research focusing on the heterogeneity of this clinical condition (e.g., Skagerlund and Träff, 2016). The findings also highlight the added value of systematically investigating primary factors during individual diagnostic research, as encouraged in the DDBP protocol. The results therefore emphasize the need for behavioral experts to investigate these factors as extensively as possible when conducting diagnostic research in severe mathematical problems. As stated in the DDBP protocol (Van Luit et al., 2014), having insight into a student's performance in these (primary) underlying factors can help experts address their problems, by giving them additional understanding of the specific educational needs of the student.

An important limitation of the current research is the absence of analysis into number sense. Research has shown that the ability to process, understand and estimate numerical quantities (Dehaene, 1992) is a predictive factor in the development of mathematical skills (Fuchs et al., 2010; LeFevre et al., 2010; De Smedt and Gilmore, 2011). Unfortunately, as noted, the lack of valid standardized tests at the start of our data collection led to the omission of this factor in this study. There have been some promising relevant testing protocols (e.g., Jordan et al., 2004), but until this moment such tests have not been standardized. In follow up research, it may be possible to investigate if deficits in this area comprise an important primary factor for DD; however, this must await a reliable and valid number sense test. Follow up research also could systematically explore the effects of secondary factors within the physical, social and educational environment. Issues such as motivation, working attitude, competence-perception and/or performance anxiety in individual students may also exert a sizable influence on performance in mathematics. These were not considered in our study.

Another limitation of the current investigation is the exclusive focus on a clinical sample. It would be desirable for follow up research to compare an atypical sample of students with a control group of students without DD. This would allow the observed

frequency of primary factors in students with DD to be compared with the occurrence of deficits in planning, naming speed, memory and attention within the normal population. Furthermore, in the current study, the clinical group originates from the client population of a single institution, which introduces the possibility of specificity. It would be useful for further research to gather broader information in order to form a more precise picture of the presence of primary and secondary factors in a more generalisable sample than in this current investigation. Nevertheless, the exploratory and descriptive nature of current research provides useful (clinical) information on systematic investigation of primary factors in students with (probable) DD.

AUTHOR CONTRIBUTIONS

JVL wrote the protocol for this research. ST analyzed the data. JVL and ST wrote the text.

FUNDING

This research has been done in the research time of both authors provided in their regular appointments in Utrecht University.

REFERENCES

- Ackerman, P. T., and Dykman, R. A. (1995). Reading-disabled students with and without comorbid arithmetic disability. *Dev. Neuropsychol.* 11, 351–371. doi: 10.1080/87565649509540625
- Alloway, T. P. (2007). *Automated Working Memory Assessment*. London: Pearson Assessment.
- Alloway, T. P., Gathercole, S. E., Kirkwood, H., and Elliott, J. (2008). Evaluating the validity of the automated working memory assessment. *Educ. Psychol.* 28, 725–734. doi: 10.1080/01443410802243828
- Alloway, T. P., Gathercole, S. E., and Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: are they separable? *Child Dev.* 77, 1698–1716. doi: 10.1111/j.1467-8624.2006.00968.x
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders*, 4th Edn. Washington, DC: Author.
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders*, 5th Edn. Washington, DC: Author. doi: 10.1176/appi.books.9780890425596
- Ashkenazi, S., and Henik, A. (2010). Attentional networks in developmental dyscalculia. *Behav. Brain Funct.* 6, 1–12. doi: 10.1186/1744-9081-6-2
- Berg, D. H. (2008). Working memory and arithmetic calculation in children: the contributory roles of processing speed, short-term memory, and reading. *J. Exp. Child Psychol.* 99, 288–308. doi: 10.1016/j.jecp.2007.12.002
- Butterworth, B., Varma, S., and Laurillard, D. (2011). Dyscalculia: from brain to education. *Science* 332, 1049–1053. doi: 10.1126/science.1201536
- Carey, E., Hill, F., Devine, A., and Szűcs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and mathematics performance. *Front. Psychol.* 6:1987. doi: 10.3389/fpsyg.2015.01987
- D'Amico, A., and Guarnera, M. (2005). Exploring working memory in children with low arithmetical achievement. *Learn. Individ. Differ.* 15, 189–202. doi: 10.1016/j.lindif.2005.01.002
- D'Amico, A., and Passolunghi, M. C. (2009). Naming speed and effortful and automatic inhibition in children with arithmetic learning disabilities. *Learn. Individ. Differ.* 19, 170–180. doi: 10.1016/j.lindif.2009.01.001
- Das, J. P., and Naglieri, J. A. (1997). *Cognitive Assessment System*. Rolling Meadows, IL: Riverside Publishing.
- De Smedt, B., and Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *J. Exp. Child Psychol.* 108, 278–292. doi: 10.1016/j.jecp.2010.09.003
- De Smedt, B., Verschaffel, L., and Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *J. Exp. Child Psychol.* 103, 469–479. doi: 10.1016/j.jecp.2009.01.010
- De Vos, T. (1992). *Handleiding Tempo-Test-Rekenen [Manual Speed Number Facts Test]*. Nijmegen: Berkhouit.
- De Vos, T. (2010). *Tempo Test Automatiseren. Handleiding en Verantwoording [Manual and Justification Speed Number Facts Test]*. Amsterdam: Boom.
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N
- Desoete, A., Roeyers, H., and De Clercq, A. (2004). Children with mathematics learning disabilities in Belgium. *J. Learn. Disabil.* 37, 50–61. doi: 10.1177/00222194040370010601
- Devine, A., Soltész, F., Nobes, A., Goswami, U., and Scücs, D. (2013). Gender differences in developmental dyscalculia depend on diagnostic criteria. *Learn. Instr.* 27, 31–39. doi: 10.1016/j.learninstruc.2013.02.004
- Dowker, A. (2005). *Individual Differences in Arithmetic. Implications for Psychology, Neuroscience and Education*. Hove: Psychology Press. doi: 10.4324/9780203324899
- Every Child a Chance Trust (2009). *The Long-Term Costs of Numeracy Difficulties*. Available at: <http://www.everychildachancetrust.org/counts/index.cfm>
- Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., et al. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Dev. Psychol.* 46, 1731–1746. doi: 10.1037/a0020662
- Fuchs, L. S., and Vaughn, S. (2012). Responsiveness-to-intervention; a decade later. *J. Learn. Disabil.* 45, 195–203. doi: 10.1177/0022219412442150
- Geary, D. C. (2004). Mathematics and learning disabilities. *J. Learn. Disabil.* 37, 4–15. doi: 10.1177/00222194040370010201

- Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *J. Dev. Behav. Pediatr.* 32, 250–263. doi: 10.1097/DBP.0b013e318209edef
- Gross-Tsur, V., Manor, O., and Shalev, R. S. (1996). Developmental dyscalculia: prevalence and demographic features. *Dev. Med. Child Neurol.* 38, 25–33. doi: 10.1111/j.1469-8749.1996.tb15029.x
- Jordan, N. C., Glutting, J., and Ramineni, C. (2004). “A number sense assessment tool for identifying children at risk for mathematical difficulties,” in *Mathematics Difficulties: Psychology and Intervention*, ed. A. Dowker (San Diego, CA: Academic Press), 45–58.
- Kaufmann, L., Mazzocco, M. M., Dowker, A., von Aster, M., Göbel, S. M., Grabner, R. H., et al. (2013). Dyscalculia from a developmental and differential perspective. *Front. Psychol.* 4:516. doi: 10.3389/fpsyg.2013.00516
- Kolkman, M. E., Kroesbergen, E. H., and Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learn. Instr.* 25, 95–103. doi: 10.1016/j.learninstruc.2012.12.001
- Koponen, T., Georgiou, G., Salmi, P., Leskinen, M., and Aro, M. (2017). A meta-analysis of the relation between RAN and mathematics. *J. Educ. Psychol.* 109, 977–992. doi: 10.1037/edu0000182
- Kroesbergen, E. H., Van Luit, J. E. H., and Naglieri, J. A. (2003). Mathematical learning difficulties and PASS cognitive processes. *J. Learn. Disabil.* 36, 574–581. doi: 10.1177/00222194030360060801
- Kroesbergen, E. H., Van Luit, J. E. H., Van der Ben, E., Leuven, N., and Vermeer, A. (2002). Meten van intelligentie bij kinderen met ADHD [Testing intelligence in children with ADHD]. *Tijdschrift voor Orthopedagogiek, Kinderpsychiatrie en Klinische Kinderpsychol.* 25, 168–179.
- Landerl, K., Bevan, A., and Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8–9-year-old students. *Cognition* 93, 99–125. doi: 10.1016/j.cognition.2003.11.004
- LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: longitudinal predictors of performance. *Child Dev.* 81, 1753–1767. doi: 10.1111/j.1467-8624.2010.01508.x
- Light, J. G., and DeFries, J. C. (1995). Comorbidity of reading and mathematics disabilities: genetic and environmental etiologies. *J. Learn. Disabil.* 23, 96–106. doi: 10.1177/002221949502800204
- Mazzocco, M. M., Feigenson, L., and Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Dev.* 82, 1224–1237. doi: 10.1111/j.1467-8624.2011.01608.x
- Mazzocco, M. M., and Grimm, K. J. (2013). Growth in rapid automatized naming from grades K to 8 in children with math or reading disabilities. *J. Learn. Disabil.* 46, 517–533. doi: 10.1177/0022219413477475
- Monuteaux, M. C., Faraone, S. V., Herzig, K., Navsaria, N., and Biederman, J. (2005). ADHD and dyscalculia: evidence for independent familial transmission. *J. Learn. Disabil.* 38, 86–93. doi: 10.1177/00222194050380010701
- Murphy, M. M., Mazzocco, M. M., Hanich, L. B., and Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *J. Learn. Disabil.* 40, 458–478. doi: 10.1177/00222194070400050901
- Mussolin, C., Mejias, S., and Noël, M. P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition* 115, 10–25. doi: 10.1016/j.cognition.2009.10.006
- Naglieri, J. A. (1999). *Essentials of CAS Assessment*. New York, NY: Wiley.
- Navarro, J. I., Aguilar, M., Alcalde, C., Ruiz, G., Marchena, E., and Menacho, I. (2011). Inhibitory processes, working memory, phonological awareness, naming speed, and early arithmetic achievement. *Span. J. Psychol.* 14, 580–588. doi: 10.5209/rev_SJOP.2011.v14.n2.6
- Olsson, L., Östergren, R., and Träff, U. (2016). Developmental dyscalculia: a deficit in the approximate number system or an access deficit? *Cogn. Dev.* 39, 154–167. doi: 10.1016/j.cogdev.2016.04.006
- Passolunghi, M. C., and Cornoldi, C. (2000). Working memory and cognitive abilities in children with specific difficulties in arithmetic word problem solving. *Adv. Learn. Behav. Disabil.* 14, 155–178.
- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012
- Raghubar, K. P., Barnes, M. A., and Hecht, S. A. (2010). Working memory and mathematics: a review of developmental, individual difference, and cognitive approaches. *Learn. Individ. Differ.* 20, 110–122. doi: 10.1016/j.lindif.2009.10.005
- Roeyers, H., and Baeyens, D. (2016). “Aandachtsprocessen [Processes of attention],” in *Handboek Diagnostiek in De leerlingbegeleiding: Kind en Context*, eds K. Verschueren and H. Koomen (Apeldoorn: Garant), 141–158.
- Rubinsten, O., and Henik, A. (2009). Developmental dyscalculia: heterogeneity might not mean different mechanisms. *Trends Cogn. Sci.* 13, 92–99. doi: 10.1016/j.tics.2008.11.002
- Ruijsenaars, A. J. J. M., Van Luit, J. E. H., and Van Lieshout, E. C. D. M. (2016). *Rekenproblemen en Dyscalculie [Arithmetic Difficulties and Dyscalculia]*. Rotterdam: Lemniscaat.
- Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20, 1–16. doi: 10.1111/desc.12372
- Shalev, R. S., Manor, O., and Gross-Tsur, V. (2005). Developmental dyscalculia: a prospective six-year follow-up. *Dev. Med. Child Neurol.* 47, 121–125. doi: 10.1017/S0012162205000216
- Skagerlund, K., and Träff, U. (2016). Number processing and heterogeneity of developmental dyscalculia: subtypes with different cognitive profiles and deficits. *J. Learn. Disabil.* 49, 36–50. doi: 10.1177/0022219414522707
- Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., et al. (2017). Math anxiety and its relationship with basic arithmetic skills among primary school children. *Br. J. Educ. Psychol.* 87, 309–327. doi: 10.1111/bjep.12151
- Toll, S. W., Van der Ven, S. H., Kroesbergen, E. H., and Van Luit, J. E. H. (2011). Executive functions as predictors of math learning disabilities. *J. Learn. Disabil.* 44, 521–532. doi: 10.1177/0022219410387302
- Träff, U., Olsson, L., Östergren, R., and Skagerlund, K. (2017). Heterogeneity of developmental dyscalculia: cases with different deficit profiles. *Front. Psychol.* 7:2000. doi: 10.3389/fpsyg.2016.02000
- Van den Bos, K. P., and Lutje Spelberg, H. C. (2007). *CB&WL. Continu Benoemen & Woorden Lezen [RN&RT. Rapid Naming & Reading Test]*. Amsterdam: Boom.
- Van Luit, J. E. H. (2012). Protocol dyscalculie als leidraad voor diagnostiek [Protocol dyscalculia as a guide for diagnostics]. *Reken-wiskundeonderwijs: Onderzoek, Ontwikkeling, Praktijk* 31, 16–21.
- Van Luit, J. E. H., Bloemert, J., Ganzinga, E. G., and Mönch, M. E. (2014). *Protocol Dyscalculie: Diagnostiek voor gedragsdeskundigen (2e herziene druk) [Protocol Dyscalculia: Diagnostics for Behaviour Professionals, 2nd Edn]*. Doetinchem: Graviant.
- Van Luit, J. E. H., Kroesbergen, E. H., and Naglieri, J. A. (2005). Utility of the PASS theory and cognitive assessment system for dutch children with and without ADHD. *J. Learn. Disabil.* 38, 434–439. doi: 10.1177/00222194050380050601

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

Copyright © 2018 Van Luit and Toll. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Cognitive and Affective Correlates of Chinese Children's Mathematical Word Problem Solving

Juan Zhang¹, Sum Kwing Cheung^{2*}, Chenggang Wu¹ and Yaxuan Meng¹

¹ Faculty of Education, University of Macau, Macau, China, ² Department of Early Childhood Education, The Education University of Hong Kong, Tai Po, Hong Kong

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

EeLynn Ng,
Nanyang Technological University,
Singapore
Annemie Desoete,
Ghent University, Belgium
Lu Wang,
Ball State University, United States

*Correspondence:

Sum Kwing Cheung
sskcheung@eduhk.hk

Specialty section:

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

Received: 04 April 2018

Accepted: 09 November 2018

Published: 18 December 2018

Citation:

Zhang J, Cheung SK, Wu C and
Meng Y (2018) Cognitive and Affective
Correlates of Chinese Children's
Mathematical Word Problem Solving.
Front. Psychol. 9:2357.
doi: 10.3389/fpsyg.2018.02357

Mathematical word problem solving (MWPS) involves multiple steps, including comprehending the problem statements, determining the arithmetic operations that have to be performed, and finding the answers. This study investigated the relative contributions of different cognitive and affective variables to children's MWPS. To achieve this goal, 116 third-grade Chinese children were tested. Results showed that after controlling for age and non-verbal intelligence, the abilities to solve direct and indirect mathematical word problems were positively correlated with the working memory component of executive function, reading comprehension ability, math fact fluency and math anxiety. Moreover, math anxiety was found to fully mediate the relationships between reading anxiety and MWPS. Implications of the findings on how to promote children's MWPS skills were discussed.

Keywords: word problems, executive function, math fact fluency, reading comprehension, math anxiety, reading anxiety, children, mathematics

INTRODUCTION

One of the major goals of mathematics learning is to know how to apply mathematical concepts to solve problems in everyday life (Kilpatrick et al., 2001). Some children, however, struggle with mathematical word problem solving (MWPS) (Hegarty et al., 1995). This happens perhaps because MWPS is not a simple task but involves at least three steps. Children have to represent the problem situation, choose a solution strategy, and apply the strategy to obtain the answer (Willis and Fuson, 1988). Therefore, MWPS does not only call for children's mathematical knowledge, but also their general cognitive skills (such as the abilities to focus only on relevant information in the problem statements, storing information of the problem situation in the working memory while retrieving possible solution strategies from the long-term memory) as well as their reading skills (Stern, 1993). Occasionally, the situation is complicated by the fact that some children possess high levels of mathematics or reading anxiety, and such negative feelings may greatly affect their performance (Tsui and Mazzocco, 2006; Wu et al., 2012; Piccolo et al., 2017; Sorvo et al., 2017). In view of the above, the present study was interested in investigating the relative contributions of different cognitive and affective variables to children's MWPS.

Word Problems: A Combination of Reading Comprehension and Mathematical Problems

Mathematical word problems refer to mathematical problems that are embedded in story contexts. Children are thus required to integrate their linguistic and basic calculation skills to find out their

solutions (Ostad, 1998). One of the many ways to classify mathematical word problems is based on the level of consistency between the language used in the story and the arithmetic operation that are called for. In direct problems, the arithmetic operation required is consistent with the relational term used in the problem (e.g., performing “addition” for a problem with the relational term “more than”) (Lewis and Mayer, 1987; Parmar et al., 1996). In contrast, in indirect problems, the arithmetic operation required is inconsistent with the relational term used in the problem (e.g., performing “addition” for a problem with the relational term “less than”) (Lewis and Mayer, 1987; Parmar et al., 1996). The present study sought to examine the correlates of children’s performance on these two types of problems because we would like to know whether the language used in mathematical word problems would affect the types of cognitive skills that were required to solve them, and the results might inform how to help those who were weak in solving different types word problems.

Several past studies have demonstrated the significant role of language and literacy skills in MWPS. Lee et al. (2009) found that children’s reading comprehension ability accounted for a significant portion of the variance in their representation of algebraic word problems. In the study of Hegarty et al. (1995), compared to unsuccessful problem solvers, successful problem solvers were less likely to adopt a direct translation strategy, i.e., simply looking for cues from numbers and keywords to come up with a plan of solving the problem. Instead, they tended to comprehend the problem and transform the problem statements into a mental representation of the problem situation (Hegarty et al., 1995). This perhaps suggests that MWPS, to certain extent, requires a deep level of reading comprehension.

On the other hand, MWPS is, no surprise, a good indicator of mathematical proficiency. In the study of Jitendra et al. (2005), children’s performance in their word problem-solving measures was positively correlated with their mathematical concepts and computational skills. Lee et al. (2009) also showed that children’s ability to discern quantitative relationships was a positive correlate of their ability to represent algebraic word problems. Interestingly, Fuchs et al. (2006) found that children’s fluency in retrieving addition and subtraction facts, but not their algorithmic computation ability, predicted their performance in arithmetic word problems.

Anxiety and Task Performance

Anxiety can impair cognitive functioning. As sub-types of anxiety, reading anxiety and math anxiety are not exceptions and are found to be linked with individuals’ performance in respective domains. Zbornik and Wallbrown (1991) found that there was a negative correlation between reading anxiety and reading achievement among upper primary school students. Similarly, Kuşdemir and Katrancı (2016) showed that fourth graders’ higher levels of reading anxiety were associated with poorer performance in a reading comprehension test. Based on a review of findings from the Program for International Student Assessment (PISA) and a number of experimental studies, Foley et al. (2017) concluded that math anxiety was negatively correlated with math performance. The relationship

was evident across countries and was likely bidirectional in nature (Foley et al., 2017). Carey et al. (2016) also supported the view of bidirectionality. As they noted, laboratory studies suggested that experimentally induced anxiety could lower individuals’ performance on math tasks, whereas data from children with dyscalculia and longitudinal studies revealed that poor math performance could evoke math anxiety (Carey et al., 2016).

To account for the mechanisms of how anxiety hampers cognitive performance, Eysenck et al. (2007)’s attentional control theory suggested that anxiety might make the individual less capable of inhibiting incorrect responses and more susceptible to distraction (e.g., threat-related stimuli that are irrelevant to the task demands, worrying thoughts). Moreover, anxiety might reduce the individual’s ability to switch attention between tasks and process the secondary task in dual-task situations (Eysenck et al., 2007).

Despite the fact that correct comprehension of the problem statements was the very first step for successful MWPS, no existing studies have examined the relative roles of reading anxiety and math anxiety in MWPS. The present study thus seeks to fill in this research gap. As demonstrated in past studies, the nature and effects of reading anxiety and math anxiety seemed to be intertwined with each other. Carey et al. (2017) found that the higher the level of math anxiety of primary and secondary school students in their sample, the poorer their mathematics as well as reading performance. In the study of Punaro and Reeve (2012), 9-year-old children who reported high levels of worries toward the language task also found the mathematical task worrying, whereas those who reported high levels of worries toward the mathematical task regarded the language task as less worrying than the mathematical task. Punaro and Reeve (2012) then concluded that literacy anxiety might be a manifestation of general academic anxiety but math anxiety was more domain-specific. Based on this theoretical notion, we speculate that math anxiety (which is a type of anxiety specifically related to the task under investigation) would mediate the relationship between reading anxiety (which is a sign of general academic anxiety) and MWPS.

Executive Function and Mathematical Learning

Executive function can be defined as the abilities to control and shift attention in a flexible manner, inhibit impulsive responses and retain information in working memory (Blair, 2016; Cantin et al., 2016). These abilities set the foundation for us to make plans, regulate emotions and control the display of impulsive acts (Blair, 2016). Different researchers have used different ways to measure executive function. Two common methods include behavioral rating scales (which can be self-rated, or rated by others like parents and teachers) and performance-based measures (e.g., Color Word Stroop task, Dimensional Change Card Sorting task) (Toplak et al., 2013; Cantin et al., 2016).

Numerous past studies have shown that children’s executive function was correlated with mathematical proficiency. Blair and Razza (2007) found that after controlling for non-verbal

intelligence, young children's inhibitory control (one component of executive function) was positively associated with their early mathematical ability. Bull and Scerif (2001) found that after controlling for intelligence and reading ability, in addition to inhibition efficiency, children's mathematical ability was also related to their perseveration and working memory span. Cantin et al. (2016) found that there was a direct positive linkage between elementary school children's cognitive flexibility and mathematical ability. Working memory, inhibitory control and cognitive flexibility also had indirect contributions to mathematical ability through reading comprehension ability (Cantin et al., 2016). In the study of Cragg et al. (2017), working memory, on the one hand, had a direct positive association with attainment in mathematics. On the other hand, the two variables were indirectly associated via knowledge of number facts, knowledge of conceptual principles underlying arithmetics and skills in performing arithmetic procedures (Cragg et al., 2017). Individuals' inhibitory control as demonstrated in a numerical task also had indirect linkages with mathematical achievement via knowledge of number facts and skills in performing arithmetic procedures (Cragg et al., 2017).

As suggested by Cawley and Miller (1986), in order to solve mathematical word problems, children have to analyze the problem situation by selecting useful information from the problem statements, followed by determining which strategy can best help solve the problem. In light of this, executive function is expected to play an important role in MWPS. For instance, Blair et al. (2015) found that there were robust associations between young children's executive function and MWPS. Swanson et al. (2008) revealed that children's accuracy in MWPS could be predicted by two components of working memory (namely central executive and visual-spatial sketchpad), as well as the growth of two components of working memory (i.e., central executive and phonological storage). Lee et al. (2009) showed that children's working memory contributed to their problem representation and solution formation for algebraic word problems. Fuchs et al. (2006) found that after controlling for non-verbal reasoning, children's attention level (as indicated by teachers' rating of their inattentive behaviors) was a positive correlate of their arithmetic word problem solving. However, working memory did not uniquely explained variance in arithmetic word problem solving, as it lost its explanatory power when phonological decoding and sight word efficiency were included in the path analysis model. Best et al. (2011) further found that children's executive function had a stronger correlation with MWPS than calculation abilities. They speculated that it was because calculation might just require the individual to retrieve mathematics facts from long-term memory, whereas MWPS often involves more plan generation and higher levels of self-monitoring (Best et al., 2011).

Even more, with emotional control as one of its components, executive function may help children regulate the negative emotions, such as anxiety, induced during the learning process. Jain and Dowson (2009), for example, found that there was a negative correlation between children's self-regulation and mathematics anxiety. Bradley et al. (2010) found that after

receiving training on emotion self-regulation, high school students in their sample showed lower levels of test anxiety. Lyons and Beilock (2012) observed the brain activities of university students with high levels of math anxiety when they were attempting a mathematical task. Based on their findings, they concluded that those who could control their cognitive resources (such as shifting attention and inhibiting predominant responses) before the task and reappraise negative emotional responses during the task tended to perform better (Lyons and Beilock, 2012). Despite the above, no existing studies have considered the contributions of math anxiety and reading anxiety when examining the relationships of executive function and domain-specific variables to MWPS. Moreover, the role of executive function in different types of MWPS has minimally been investigated. Compared to direct MWPS, we speculate that indirect MWPS require more executive function resources to solve, because children have to inhibit their intuitive response of carrying out arithmetic operations simply based on the relational term given (Daroczy et al., 2015). They also have to be more cognitively flexible in order to rephrase the inconsistent relational sentence and represent the problems properly (Lewis and Mayer, 1987).

Present Study

As discussed, MWPS is a crucial part of mathematics learning. Despite the attention received by various researchers, the contributions of different cognitive skills (including general and domain-specific ones) to children's MWPS have seldom been compared, and the potential role of affective variables in children's MWPS has often been overlooked. The present study thus sought to examine Chinese children's MWPS in relation to an array of cognitive and affective variables.

The cognitive variables under investigation included: (1) non-verbal intelligence, (2) executive function (including five components, namely inhibit, shift, emotional control, working memory, and plan/organize), (3) math fact fluency, and (4) reading comprehension. The first two were selected because these general cognitive skills play an important role in many types of cognitive processing (Fuchs et al., 2006; Jain and Dowson, 2009; Lee et al., 2009). The third was selected because good foundation skills in mathematics might facilitate children to solve higher-level mathematical problems (Binder, 1996; McCallum et al., 2006). The last was selected because past studies have found a close linkage between children's literacy and mathematical development (Purpura et al., 2011), and MWPS required children to represent the problem embedded in a piece of text mathematically (Hegarty et al., 1995). Meanwhile, the affective variables of interest were: (1) math anxiety and (2) reading anxiety. These two variables are selected because it is not uncommon for children to have these kinds of anxiety, and these negative feelings have often been to hinder children's performance in related tasks (Piccolo et al., 2017; Sorvo et al., 2017).

To obtain a more comprehensive capture of children's MWPS skills, their performance was assessed with two tasks, namely direct and indirect problems. Compared to direct problems, indirect problems may be more challenging. This

is because children have to be careful not to misinterpret the relational statements and perform arithmetic operations that are inconsistent with the relational term used in the problem statements (Lewis and Mayer, 1987; Daroczy et al., 2015). The whole indirect MWPS process may thus call for more and a wider range of cognitive resources.

Behavioral rating scales would be used to measure executive function, math anxiety and reading anxiety, whereas performance-based measures would be adopted to assess the remaining variables. We rely on behavioral rating scale rather than performance-based measure for executive function because two of the variables under focus were related to anxiety. We thus wanted to assess executive function skills as displayed in natural everyday life rather than stressful test situation. The behavioral rating scale could allow us to assess the “emotional control” component of executive function.

Based on results of past studies (e.g., Fuchs et al., 2006; Lee et al., 2009; Punaro and Reeve, 2012; Cantin et al., 2016; Carey et al., 2017), the followings are hypothesized:

H1a: Children’s MWPS is positively related to their non-verbal intelligence, executive function, math fact fluency, and reading comprehension ability.

H1b: Children’s MWPS is negatively related to their math anxiety and reading anxiety.

H2: Compared to direct MWPS, indirect MWPS has stronger correlations with executive function and reading comprehension ability.

H3: Children’s math anxiety mediated the relationship between their reading anxiety and MWPS.

To test our hypotheses, six-step hierarchical linear regression analyses would be performed on direct and indirect MWPS, respectively. In the first step, age and non-verbal IQ would be entered, so as to control their effects on direct and indirect MWPS. In the second step, executive function would be entered because this could allow us to understand the extent to which the contribution of the domain-general variable (i.e., executive function) to MWPS was shared by the domain-specific ones. Instead of combining the five components of executive function into one latent variable and entering it into the regression equation, the five components would be entered in stepwise fashion. This was important because indirect MWPS might require higher levels of inhibition of prepotent response than direct MWPS, and such analytic approach could help us know whether different component(s) of executive function had differential associations with direct and indirect MWPS. In the third step, math fact fluency would be entered because it assessed one’s basic mathematical abilities and was apparently the most relevant to MWPS. Adding it into the regression equations at this stage could let us examine at later stages whether other domain-specific cognitive and affective variables could make unique contributions to MWPS after controlling one’s basic mathematical abilities. The remaining three variables (i.e., math anxiety, reading anxiety, and reading comprehension ability) would be entered alternatively in the fourth to sixth steps. This allowed us to investigate whether each of them could make

unique contributions to children’s MWPS. To test Hypothesis 3, Baron and Kenny’s (1986) procedures would be used, as it has extensively been used for mediation analysis among past studies, including recently published ones (e.g., Zhang et al., 2017; Kam et al., 2018; Mazzone et al., 2018).

MATERIALS AND METHODS

Participants and Procedure

The participants were 116 third-grade primary students (61 boys and 55 girls, mean age = 9.60 years, $SD = 0.50$) from Zhuhai, Guangdong Province of China. All participants were typically developing children and written consent forms were collected from their parents or guardians prior to the formal tests. The experimental procedures were approved by the Ethics Committee of University of Macau. All tests were carried out in accordance with the approved guidelines and regulations. Six tasks were administered to each child through two sessions, including non-verbal intelligence, reading comprehension, reading anxiety (session 1), executive function, math fact fluency, MWPS, and math anxiety (session 2). The interval between the two sessions was about 1 week and each session lasted for about 2 h. Children could ask for a rest during the assessment. The executive function of children was evaluated by their parents using the BRIEF scale.

Measures

Non-verbal Intelligence

Non-verbal intelligence was assessed using the Raven’s Progressive Matrix (Set A and Set B) (Kratzmeier and Horn, 1980). Specifically, two sets were used with each contained 12 items. During the assessment, the child was tested independently by research assistants and was asked to choose one piece of figure that could best fit the missing part of a visual geometric picture from six options. Each correct answer was scored as 1 and the maximum total score was 24 (Cronbach’s $\alpha = 0.79$).

Executive Function

The Behavior Rating Inventory of Executive Function (BRIEF) (Gioia et al., 2000) was used, because it was a standardized questionnaire and had widely been used to assess the executive function (EF) of children aged from 5 to 18 years old among past studies (e.g., Anderson, 2002; Isquith et al., 2004; Roth et al., 2014). There were two versions of the BRIEF scale (teacher version and parent version). In the present study, only the parent version was adopted because parents were likely to be more familiar with children’s performance in everyday life. Five dimensions of children’s EF, including inhibit (10 items), emotional control (10 items), shift (8 items), working memory (10 items) and plan/organize (12 items), were assessed. For each item, parents were asked to rate on a three-point scale. The higher the children were scored, the worse their EF was (Cronbach’s $\alpha = 0.83$ for inhibit, Cronbach’s $\alpha = 0.73$ for shift, Cronbach’s $\alpha = 0.73$ for emotional control, Cronbach’s $\alpha = 0.85$ for working memory, and Cronbach’s $\alpha = 0.78$ for plan/organization).

Reading Comprehension

During the task (Zhang et al., 2014), children were asked to read each of the given sentences silently and then choose one picture that best fits the sentence in meaning. There were 30 items, and their difficulty levels were appropriate for third-grade primary children in China. The maximum score was 30 (Cronbach's $\alpha = 0.67$).

Math Fact Fluency

This test was adopted and modified from the tasks used by Fuchs et al. (2006) and Nosworthy et al. (2013). In this test, 50 addition, 50 subtraction, and 50 multiplication items were presented to students, respectively, and students were asked to finish answering each type of items in 1 min. Before the test, the test paper was sent out to students with the back of the paper upward and no student was permitted to answer the questions until the research assistant told them to start. All answers should be provided within 1 min and students should stop answering when the time was over. Each correct answer was scored as 1 and the maximum score for the test was 150 (Cronbach's $\alpha = 0.89$).

Math Word Problem Solving

With reference to related past studies (Lewis and Mayer, 1987; Nesher and Hershkovitz, 1994; Parmar et al., 1996), 22 two-step mathematical word problems were created in accordance with the mathematical abilities and social experience of third-grade students in China. Of the two steps, one step involved addition or subtraction, and the other step involved multiplication or division. Half of the problems were direct problems, in which the relational term used in the problem was consistent with the arithmetic operation required for one step of the problem (e.g., A diary book costs \$23. Uncle Ho bought 72 diary books. Uncle Wong bought 25 diary books fewer than Uncle Ho. How much should Uncle Wong pay?). The remaining half was indirect problems, in which the relational term used in the problem was inconsistent with the arithmetic operation required for one step of the problem (e.g., An exercise book costs \$23. Uncle Yeung bought 72 exercise books. Uncle Yeung bought 28 exercise books more than Uncle Lee. How much should Uncle Lee pay?). The presentation order of the two kinds of problems was counterbalanced among participants, and participants were asked to solve all the problems within a limited period of time. The maximum score for each problem type was 11 (Cronbach's $\alpha = 0.81$ for direct problems and 0.74 for indirect problems).

Math Anxiety

The Test Anxiety Scale of Pintrich et al. (1991) was adopted. The original items about general test anxiety were modified to make them specific to mathematics (e.g., changing the word “test” to the phrase “math test”). Five 7-point items, ranging from 1 (totally incorrect) to 7 (totally correct), were used to evaluate the math anxiety. The participants were asked to choose the number of degree that best fit their status toward math. A sample item was “I have an uneasy, upset feeling when I take a math exam.” The maximum score for the scale was 35 (Cronbach's $\alpha = 0.76$).

Reading Anxiety

This scale consisted of five items. The items and instructions given to participants were exactly the same as those of the math anxiety scale, except that the word “math” was replaced by the word “reading.” The maximum score was 35 (Cronbach's $\alpha = 0.89$).

RESULTS

The descriptive summary of all the variables is displayed in **Table 1**. As expected, the mean score of indirect MWPS (3.11) was significantly lower than that of direct MWPS (3.78), $t(115) = 4.81$, $p < 0.001$, showing that indirect word problems were harder than direct word problems.

Table 2 shows the correlations and partial correlations (with age and non-verbal IQ controlled) among all the variables. As illustrated in **Table 2**, the construct validity of executive function measurement was confirmed by the strong associations among the four dimensions (i.e., inhibit, shift, emotion control, and working memory), except the “plan/organize” dimension. Both the zero-order correlation and partial correlation analyses suggested that two aspects of executive function (i.e., inhibit and working memory) strongly correlated with direct and indirect MWPS. As expected, math fact fluency and math anxiety were significantly correlated with both direct and indirect MWPS. Reading comprehension and reading anxiety were also strongly associated with direct and indirect MWPS.

To further examine the contributions of executive function, math fact fluency, reading anxiety, math anxiety, and reading comprehension to direct MWPS and indirect MWPS, six-step hierarchical regressions were conducted (see **Tables 3, 4**). As described earlier, age and non-verbal IQ were entered as control variables in the first step. In the second step, the five components of executive function were entered by stepwise. Only working memory remained in the model and accounted another 9% variance for direct MWPS and 8% variance for indirect MWPS. Math fact fluency was entered in the third

TABLE 1 | Descriptive statistics for all measurements.

	Minimum	Maximum	Mean	SD
Non-verbal IQ	12.00	24.00	20.99	2.56
EF – Inhibit	37.00	69.00	49.96	7.94
EF – Shift	36.00	74.00	50.32	9.24
EF – Emotional control	36.00	67.00	48.92	8.20
EF – Working memory	38.00	74.00	52.68	8.24
EF – Plan/Organize	33.00	58.00	51.66	5.33
Math fact fluency	42.00	150.00	97.37	24.19
Direct MWPS	0.00	9.00	3.78	2.42
Indirect MWPS	0.00	10.00	3.11	2.20
Math anxiety	1.00	7.00	2.82	1.68
Reading comprehension	4.00	28.00	20.99	4.53
Reading anxiety	1.00	7.00	3.07	1.74

The scores of the EF dimensions had been transformed to standard scores according to the BRIEF Manual.

TABLE 2 | Correlations and partial correlations among variables.

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Age	1.00												
2. Non-verbal IQ	−0.10	1.00											
3. EF – Inhibit	−0.01	−0.09	1.00	0.60***	0.64***	0.74***	0.01	−0.16	−0.26**	−0.24**	0.12	−0.13	0.21*
4. EF – Shift	−0.18	−0.10	0.60***	1.00	0.70***	0.59***	0.08	−0.17	−0.13	−0.18	0.14	−0.09	0.17
5. EF – Emotional control	−0.13	0.08	0.62***	0.69***	1.00	0.59***	0.02	−0.07	−0.09	−0.09	0.05	−0.12	0.09
6. EF – Working memory	0.10	−0.07	0.74***	0.60***	0.59***	1.00	0.07	−0.13	−0.31**	−0.29**	0.07	−0.17	0.15
7. EF – Plan/Organize	−0.51**	−0.07	0.02	0.17	0.07	0.12	1.00	0.16	−0.04	−0.08	−0.08	0.04	0.07
8. Math fact fluency	−0.31**	0.05	−0.15	−0.10	−0.03	−0.10	−0.28**	1.00	0.46***	0.44***	−0.15	0.18	−0.08
9. Direct MWPS	−0.12	0.12	−0.27**	−0.12	−0.07	−0.30**	−0.01	0.47**	1.00	0.79***	−0.40***	0.35***	−0.30**
10. Indirect MWPS	−0.22*	0.19*	−0.25**	−0.15	−0.04	−0.27**	−0.03	0.48**	0.80**	1.00	−0.39***	0.33***	−0.32**
11. Math anxiety	0.08	−0.13	0.130	0.14	0.04	0.07	−0.09	−0.17	−0.41**	−0.41**	1.00	−0.18	0.77***
12. Reading comprehension	−0.08	0.21*	−0.15	−0.09	−0.09	−0.17	0.06	0.20*	0.36**	0.36**	−0.20*	1.00	−0.16
13. Reading anxiety	−0.02	−0.10	0.21*	0.18	0.08	0.16	0.09	−0.08	−0.30**	−0.32**	0.77**	−0.17*	1.00

Correlations among variables are presented below the diagonal; partial correlations controlling for age and non-verbal intelligence are presented above the diagonal. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

step and could account for another 17% variance for direct MWPS and 15% variance for indirect MWPS. Reading anxiety could account for extra 5% variance for direct MWPS and 6% variance for indirect MWPS when age, non-verbal IQ, working memory, and math fact fluency were statistically controlled. In addition, math anxiety could uniquely predict both direct (5%) and indirect (3%) MWPS in step 5. More importantly, reading comprehension could still account for 3% variance for both direct and indirect MWPS in step 6, indicating a unique role of reading comprehension in solving math word problems. We performed another set of hierarchical regressions to investigate the role of reading comprehension in the relationship between reading anxiety and MWPS. Results showed that reading anxiety could still account for 4% variance for direct and indirect MWPS respectively after reading comprehension was controlled. Finally, with a purpose to explore the role of math anxiety in the relationship between reading anxiety and MWPS, another set of hierarchical regressions

TABLE 4 | Hierarchical regression predicting indirect math word problem solving.

Step	β	T	R^2	ΔR^2
1 Age	−0.21	−2.27*	0.08	0.08*
Non-verbal IQ	0.17	1.86		
2 EF – Working memory	−0.28	−3.24**	0.16	0.08**
3 Math fact fluency	0.42	4.99***	0.31	0.15***
4 Reading anxiety	−0.25	−3.23**	0.37	0.06**
5 Math anxiety	−0.29	−2.44**	0.40	0.03*
6 Reading comprehension	0.18	2.26*	0.43	0.03*
4 Reading comprehension	0.22	2.73**	0.36	0.05**
5 Reading anxiety	−0.23	−2.96**	0.40	0.04**
6 Math anxiety	−0.27	−2.29*	0.43	0.03*
4 Reading comprehension	0.22	2.73**	0.36	0.05**
5 Math anxiety	−0.28	−3.80***	0.43	0.07***
6 Reading anxiety	−0.02	−0.17	0.43	0

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

TABLE 3 | Hierarchical regression predicting direct math word problem solving.

Step	β	T	R^2	ΔR^2
1 Age	−0.11	−1.16	0.03	0.03
Non-verbal IQ	0.11	1.19		
2 EF – Working memory	−0.31	−3.419**	0.12	0.09**
3 Math fact fluency	0.44	5.22***	0.29	0.17***
4 Reading anxiety	−0.23	−2.92**	0.34	0.05**
5 Math anxiety	−0.37	−3.07**	0.39	0.05**
6 Reading comprehension	0.19	2.49*	0.42	0.03*
4 Reading comprehension	0.24	2.95**	0.34	0.05**
5 Reading anxiety	−0.20	−2.63*	0.38	0.04*
6 Math anxiety	−0.34	−2.91*	0.43	0.05**
4 Reading comprehension	0.24	2.95**	0.34	0.05**
5 Math anxiety	−0.30	−3.97***	0.42	0.08**
6 Reading anxiety	0.06	0.50	0.42	0

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

was conducted. Surprisingly, after math anxiety was entered in step 5, reading anxiety could not predict direct or indirect MWPS in step 6, suggesting that the relationship between reading anxiety and MWPS was possibly mediated by math anxiety.

To provide more direct evidence confirming that reading anxiety was associated with direct MWPS and indirect MWPS with the mediation of math anxiety, we combined direct MWPS and indirect MWPS as math word problem solving (WPS) and conducted additional hierarchical regression analysis. Four conditions should be met to confirm the full mediation effect (Baron and Kenny, 1986). First, the reading anxiety (predictor) should be significantly regressed (c) on math WPS (outcome). Second, the mediator (math anxiety) should also be associated with the outcome (b). Third, the predictor and mediator should be closely related (a). Fourth, the predictor should be non-significant after mediator was controlled ($c' < c$). As illustrated in Table 5, reading anxiety (condition 1) and math anxiety

(condition 2) could both predict math WPS (direct and indirect MWPS combined), respectively. In addition, reading anxiety was significantly associated with math anxiety (condition 3). More importantly, reading anxiety failed to predict math WPS after math anxiety was statistically controlled. Therefore, the result showed that reading anxiety was fully mediated by math anxiety when predicting math WPS (see Table 5 and Figure 1).

DISCUSSION

The present study examined the relative contributions of different cognitive and affective variables to children’s MWPS. Our findings suggested that after controlling for age and non-verbal IQ, children’s MWPS (no matter for direct or indirect problems) was only significantly correlated with the working memory component of executive function, math fact fluency, reading comprehension and math anxiety. Regarding the relationships of executive function and reading comprehension to MWPS, their strengths were similar across direct and indirect problems. Moreover, the association between reading anxiety and MWPS was fully mediated by math anxiety.

Partially different from our initial speculations, only two of the five components of executive function (i.e., inhibit and working memory) had significant but weak zero-order correlations with MWPS (including direct and indirect problems). No significant associations were found for the remaining three components (i.e., shift, emotional control, and plan/organize). However, when age, non-verbal intelligence and the five components

were considered together in the regression equations, only working memory remained as a significant correlate. This is somehow similar to the findings by Lee et al. (2009), in which only working memory, but not inhibition and mental flexibility, were significantly related to skills of solving algebraic problems. Moreover, different from our hypothesis, the strength of the relationship between executive function and MWPS did not vary much across direct and indirect problems. These findings perhaps suggest that working memory shares some overlapping roles with other components in MWPS but its role is relatively more prominent. This happens possibly because during MWPS, children have to handle multiple tasks within a fairly short period of time, such as comprehending the problem statements, memorizing various pieces of useful information about the problem situation, forming a mental representation of the problem situation in mathematical terms, recalling possible solution strategies (Willis and Fuson, 1988). With better working memory, children can perform the aforesaid tasks more efficiently (Lee et al., 2009). Meanwhile, compared to working memory, other components of executive function might be relatively less crucial, given that the mathematical word problems involved were not very complicated in nature (i.e., two-step word problems only).

Consistent with our initial speculation and results of past studies (Fuchs et al., 2006), math fact fluency was a significant correlate of children’s direct and indirect MWPS. To recall, the math fact fluency task required children to retrieve basic addition, subtraction and multiplication facts accurately and quickly. The ability to perform such a task well might thus help children to

TABLE 5 | Hierarchical regression examining the mediating effect of math anxiety.

	Condition 1		Condition 2		Condition 3		Condition 4	
	β	t	β	t	β	t	β	T
Reading anxiety	−0.33	−3.69***	–	–	0.77	12.94***	0.02	0.12
Math anxiety	–	–	−0.43	−5.11***			−0.44	−3.33***

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

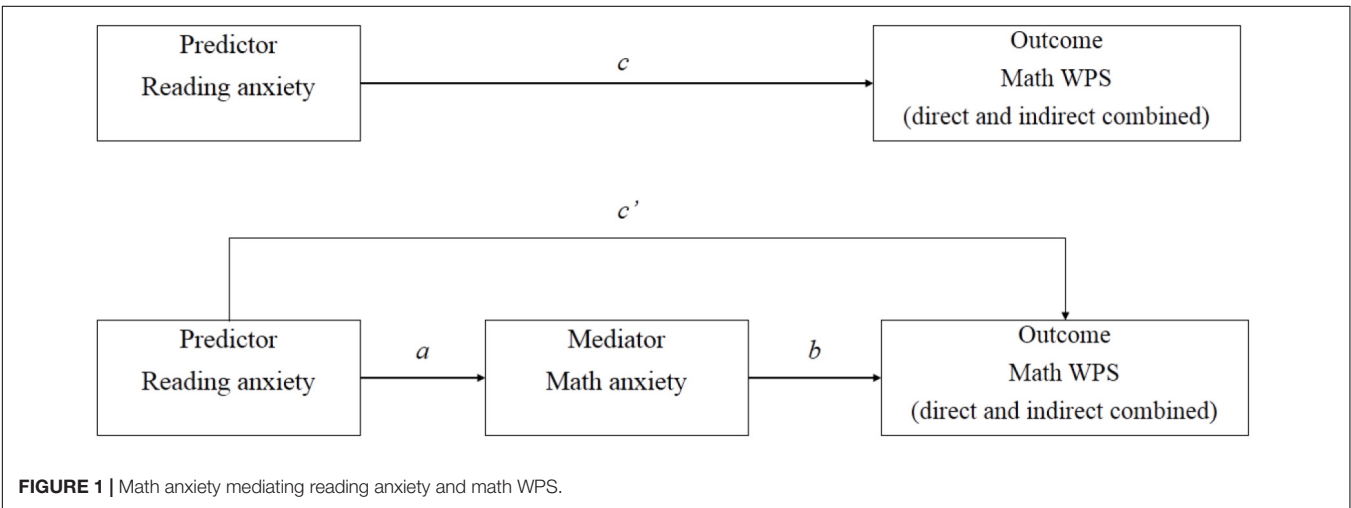


FIGURE 1 | Math anxiety mediating reading anxiety and math WPS.

free up more mental resources to handle more complicated tasks required in the process of MWPS (Wong, 1986).

As expected, reading comprehension was a significant correlate of both types of MWPS. This again shows that it is important for children to have a deep comprehension of the problem statements, instead of just relying on the keywords in the problem statements (e.g., the number words, the relational terms) to solve the mathematical word problems (Hegarty et al., 1995). However, different from our hypothesis, the strength of its relationship with MWPS did not differ much across direct and indirect problems. This happens perhaps because inconsistent language only appeared in one of the several problem statements in indirect problems. Thus, compared to direct problems, solving indirect problems might not call for much extra reading comprehension skills.

Of the two affective variables under investigation, math anxiety seemed to be more relevant to MWPS than reading anxiety was. On the one hand, though reading anxiety had significant zero-order correlations with both types of MWPS, it was not a significant correlate when math anxiety was included in the regression equations. On the other hand, even after controlling for all the cognitive variables, math anxiety could still account for additional variance in both types of MWPS. These results, somehow, are not surprising for at least two reasons. First, as proposed by Punaro and Reeve (2012), reading anxiety can be regarded as a sign of general academic anxiety. It may therefore lose its power to explain variations in children's MWPS when it is considered together with math anxiety (i.e., a specific type of anxiety associated with the academic domain under examination). Second, reading comprehension is only the very first step of MWPS, and literal understanding the problem situation does not necessarily lead to a correct answer. The fear of failure in representing the problem situation from a mathematical perspective and performing the required arithmetic operation might thus be more closely related to the outcome of MWPS, i.e., obtaining the correct answer. Indeed, high level of math anxiety can hinder children's performance on mathematical tasks by creating a disruption of their working memory (Ashcraft and Kirk, 2001). Meanwhile, it is also possible that children with high level of math anxiety might be less willing to engage in math-related tasks in their everyday life, which in turn make them have fewer opportunities to practice their math skills and develop competency in MWPS (Fennema, 1989). Nevertheless, it should be noted that compared to the possible score range (i.e., 1–7), the mean scores of math anxiety and reading anxiety of our participants were only 2.82 and 3.07. In fact, in samples with higher levels of math anxiety or reading anxiety, the contributions of executive function and anxiety variables might become even more crucial, because individual might suffer from greater impairments on working memory and other executive function components (e.g., inhibit, plan/organize) and it might be more important to possess higher levels of the “emotional control” component for maintaining performance.

Findings of the present study can provide educators and parents with insights on how to promote children's MWPS

skills. First, when teachers observe that a child makes mistakes frequently when solving mathematical word problems, teachers have to examine more closely the reason(s) for such a situation. As shown in the present study, it might happen because the child shows difficulties in processing multiple pieces of information in the mind, comprehending the problem statements, and/or retrieving basic math facts to find out the answers. These different reasons indeed call for different approaches to help the child.

Second, the present study shows that teachers have to find out effective strategies to help children reduce math anxiety. This is because the fear induced by the necessity of tackling math problems might be so overwhelming that it can hinder children's math performance, even though the children might have already possessed the required math knowledge and skills. Teachers and parents should thus talk to children who show high levels of anxiety, so as to understand the reasons underlying their anxiety and adopt corresponding strategies to relieve their stress.

The present study had several limitations that required attention. First, given that all variables were measured at one time point only, no causal relationships between the variables can be drawn. Future researchers can thus conduct longitudinal studies to examine the extent to which various cognitive and affective variables and their growth can predict children's performance in MWPS in future.

Second, the present study only relied on parental report questionnaire to measure children's executive function. Some recent studies (e.g., Toplak et al., 2008, 2013) have suggested that there were only modest correlations between the scores obtained from behavioral rating scales and performance-based measures and the two types of measures might assess two different cognitive levels. Performance-based measures might assess the efficiency of the cognitive processes that are employed for controlling behaviors, whereas behavioral rating scales might tap on the behaviors of how to achieve personal goals in real-life situations. In the future, researchers can therefore measure executive function using different methods and examined whether the patterns of results found in the present study still hold the same.

Third, the present study only focused on two types of word problems (i.e., direct and indirect two-step arithmetic word problems) and the outcome of children's MWPS (i.e., the accuracy of their answers). In fact, the difficulty level of arithmetic word problems depends on a range of linguistic, numerical and contextual factors (Cummins et al., 1988; Davis-Dorsey et al., 1991; Daroczy et al., 2015), and there are different correlates of children's performance in different cognitive phases of MWPS (Lee et al., 2009). Future studies can thus explore whether the variables associated with each MWPS stage are the same across problem types and test situations. This, in turn, can yield important implications on how to help children tackle different types of word problems.

In summary, the present study is one of the few studies to investigate the relative contributions of different cognitive

and affective variables to children's MWPS. Our findings showed that children's MWPS was significantly correlated with their working memory, reading comprehension, math fact fluency and math anxiety. In order to provide more effective support to children struggling with MWPS, it is essential for teachers and parents to figure out in which aspect these children show difficulties and adopt corresponding strategies to help them overcome their barriers.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of ethic guidelines, Institutional Review Board in the University of Macau with written informed consent from all subjects. The protocol was approved by the Institutional Review Board in the University of Macau.

REFERENCES

- Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychol.* 8, 71–82. doi: 10.1076/chin.8.2.71.8724
- Ashcraft, M. H., and Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *J. Exp. Psychol. Gen.* 130, 224–237. doi: 10.1037/0096-3445.130.2.224
- Baron, R. M., and Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51, 1173–1182. doi: 10.1037/0022-3514.51.6.1173
- Best, J. R., Miller, P. H., and Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learn. Individ. Differ.* 21, 327–336. doi: 10.1016/j.lindif.2011.01.007
- Binder, C. (1996). Behavioral fluency: evolution of a new paradigm. *Behav. Analyst.* 19, 163–197. doi: 10.1007/BF03393163
- Blair, C. (2016). Developmental science and executive function. *Curr. Dir. Psychol. Sci.* 25, 3–7. doi: 10.1177/0963721415622634
- Blair, C., and Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Dev.* 78, 647–663. doi: 10.1111/j.1467-8624.2007.01019.x
- Blair, C., Ursache, A., Greenberg, M., and Vernon-Feagans, L. (2015). Multiple aspects of self-regulation uniquely predict mathematics but not letter-word knowledge in the early elementary grades. *Dev. Psychol.* 51, 459–472. doi: 10.1037/a0038813
- Bradley, R. T., McCraty, R., Atkinson, M., Tomasino, D., Daugherty, A., and Arguelles, L. (2010). Emotion self-regulation, psychophysiological coherence, and test anxiety: results from an experiment using electrophysiological measures. *Appl. Psychophysiol. Biofeedback* 35, 261–283. doi: 10.1007/s10484-010-9134-x
- Bull, R., and Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: inhibition, switching, and working memory. *Dev. Neuropsychol.* 19, 273–293. doi: 10.1207/S15326942DN1903-3
- Cantin, R. H., Gnaedinger, E. K., Gallaway, K. C., Hesson-McInnis, M. S., and Hund, A. M. (2016). Executive functioning predicts reading, mathematics, and theory of mind during the elementary years. *J. Exp. Child Psychol.* 146, 66–78. doi: 10.1016/j.jecp.2016.01.014
- Carey, E., Devine, A., Hill, F., and Szűcs, D. (2017). Differentiating anxiety forms and their role in academic performance from primary to secondary school. *PLoS One* 12:e0174418. doi: 10.1371/journal.pone.0174418
- Carey, E., Hill, F., Devine, A., and Szűcs, D. (2016). The chicken or the egg? The direction of the relationship between mathematics anxiety and

AUTHOR CONTRIBUTIONS

JZ and SKC developed the research idea and research design. CW analyzed the data. JZ, SKC, CW, and YM wrote and reviewed the manuscript.

FUNDING

This study was supported by MYRG2017-00217-FED, MYRG2016-00193-FED, and MYRG2015-00221-FED from the University of Macau in Macau.

ACKNOWLEDGMENTS

We would like to thank all the children and parents who had participated in our study.

- mathematics performance. *Front. Psychol.* 6:1987. doi: 10.3389/fpsyg.2015.01987
- Cawley, J. F., and Miller, J. H. (1986). Selected views on metacognition, arithmetic problem solving, and learning disabilities. *Learn. Disabil. Focus* 2, 36–48.
- Cragg, L., Keeble, S., Richardson, S., Roome, H. E., and Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition* 162, 12–26. doi: 10.1016/j.cognition.2017.01.014
- Cummins, D. D., Kintsch, W., Reusser, K., and Weimer, R. (1988). The role of understanding in solving word problems. *Cognit. Psychol.* 20, 405–438. doi: 10.1016/0010-0285(88)90011-4
- Daroczy, G., Wolska, M., Meurers, W. D., and Nuerk, H. C. (2015). Word problems: a review of linguistic and numerical factors contributing to their difficulty. *Front. Psychol.* 6:348. doi: 10.3389/fpsyg.2015.00348
- Davis-Dorsey, J., Ross, S. M., and Morrison, G. R. (1991). The role of rewording and context personalization in the solving of mathematical word problems. *J. Educ. Psychol.* 83, 61–68. doi: 10.1037/0022-0663.83.1.61
- Eysenck, M. W., Derakshan, N., Santos, R., and Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion* 7, 336–353. doi: 10.1037/1528-3542.7.2.336
- Fennema, E. (1989). "The study of affect and mathematics: a proposed generic model for research," in *Affect and Mathematical Problem Solving*, eds D. B. McLeod and V. M. Adams (New York, NY: Springer), 205–219. doi: 10.1007/978-1-4612-3614-6_14
- Foley, A. E., Herts, J. B., Borgonovi, F., Guerriero, S., Levine, S. C., and Beilock, S. L. (2017). The math anxiety-performance link: a global phenomenon. *Curr. Dir. Psychol. Sci.* 26, 52–58. doi: 10.1177/0963721416672463
- Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *J. Educ. Psychol.* 98, 29–43. doi: 10.1037/0022-0663.98.1.29
- Gioia, G. A., Isquith, P. K., Guy, S. C., and Kenworthy, L. (2000). *The Behavior Rating Inventory of Executive Function (BRIEF)*. Odessa, FL: Psychological Assessment Resources.
- Hegarty, M., Mayer, R. E., and Monk, C. A. (1995). Comprehension of arithmetic word problems: a comparison of successful and unsuccessful problem solvers. *J. Educ. Psychol.* 87, 18–32. doi: 10.1037/0022-0663.87.1.18
- Isquith, P. K., Gioia, G. A., and Espy, K. A. (2004). Executive function in preschool children: examination through everyday behavior. *Dev. Neuropsychol.* 26, 403–422. doi: 10.1207/s15326942dn2601_3
- Jain, S., and Dowson, M. (2009). Mathematics anxiety as a function of multidimensional self-regulation and self-efficacy. *Contemp. Educ. Psychol.* 34, 240–249. doi: 10.1016/j.cedpsych.2009.05.004

- Jitendra, A. K., Sczesniak, E., and Deatline-Buchman, A. (2005). An exploratory validation of curriculum-based mathematical word problem-solving tasks as indicators of mathematics proficiency for third graders. *School Psych. Rev.* 34, 358–371.
- Kam, C. C. S., Hue, M. T., and Cheung, H. Y. (2018). Academic dishonesty among Hong Kong secondary school students: application of theory of planned behaviour. *Educ. Psychol.* 38, 945–963. doi: 10.1080/01443410.2018.1454588
- Kilpatrick, J., Swafford, J., and Findell, B. (2001). “The strands of mathematical proficiency,” in *Adding It Up: Helping Children Learn Mathematics*, eds J. Kilpatrick, J. Swafford, and B. Findell (Washington: National Academy Press), 115–118.
- Kratzmeier, H., and Horn, R. (1980). *APM-Advanced Progressive Matrices (Raven-Matrizen-Test)*. Weinheim: Beltz.
- Kuşdemir, Y., and Katrancı, M. (2016). Anxiety and comprehension in reading: i cannot find the main idea, my teacher! *Educ. Sci.* 41, 251–266. doi: 10.15390/EB.2016.4951
- Lee, K., Ng, E. L., and Ng, S. F. (2009). The contributions of working memory and executive functioning to problem representation and solution generation in algebraic word problems. *J. Educ. Psychol.* 101, 373–387. doi: 10.1037/a0013843
- Lewis, A. B., and Mayer, R. E. (1987). Students’ miscomprehension of relational statements in arithmetic word problems. *J. Educ. Psychol.* 79, 363–371. doi: 10.1037/0022-0663.79.4.363
- Lyons, I. M., and Beilock, S. L. (2012). Mathematics anxiety: separating the math from the anxiety. *Cereb. Cortex* 22, 2102–2110. doi: 10.1093/cercor/bhr289
- Mazzone, A., Camodeca, M., and Salmivalli, C. (2018). Stability and change of outsider behavior in school bullying: the role of shame and guilt in a longitudinal perspective. *J. Early Adolesc.* 38, 164–177. doi: 10.1177/0272431616659560
- McCallum, E., Skinner, C. H., Turner, H., and Lee, S. (2006). The taped-problems intervention: increasing multiplication fact fluency using a low-tech, classwide, time-delay intervention. *School Psych. Rev.* 35, 419–434.
- Nesher, P., and Hershkovitz, S. (1994). The role of schemes in two-step problems: analysis and research findings. *Educ. Stud. Math.* 26, 1–23. doi: 10.1007/BF01273298
- Nosworthy, N., Bugden, S., Archibald, L., Evans, B., and Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children’s arithmetic competence. *PLoS One* 8:e67918. doi: 10.1371/journal.pone.0067918
- Ostad, S. A. (1998). Developmental differences in solving simple arithmetic word problems and simple number-fact problems: a comparison of mathematically normal and mathematically disabled children. *Math. Cogni.* 4, 1–19. doi: 10.1080/135467998387389
- Parmar, R. S., Cawley, J. F., and Frazita, R. R. (1996). Word problem-solving by students with and without mild disabilities. *Except. Child.* 62, 415–429. doi: 10.1177/001440299606200503
- Piccolo, L. R., Giacomoni, C. H., Julio-Costa, A., Oliveira, S., Zbornik, J., Haase, V. G., et al. (2017). Reading anxiety in L1: reviewing the concept. *Early Child. Educ. J.* 45, 537–543. doi: 10.1007/s10643-016-0822-x
- Pintrich, P., Smith, D., Garcia, T., and McKeachie, W. (1991). *A Manual for the Use of The Motivated Strategies for Learning Questionnaire (MSLQ)*. Ann Arbor, MI: National Center for Research to Improve Postsecondary Teaching and Learning.
- Punaro, L., and Reeve, R. (2012). Relationships between 9-year-olds’ math and literacy worries and academic abilities. *Child Dev. Res.* 2012, 1–11. doi: 10.1155/2012/359089
- Purpura, D. J., Hume, L. E., Sims, D. M., and Lonigan, C. J. (2011). Early literacy and early numeracy: the value of including early literacy skills in the prediction of numeracy development. *J. Exp. Child Psychol.* 110, 647–658. doi: 10.1016/j.jecp.2011.07.004
- Roth, R. M., Isquith, P. K., and Gioia, G. A. (2014). “Assessment of executive functioning using the Behavior Rating Inventory of Executive Function (BRIEF),” in *Handbook of Executive Functioning*, eds S. Goldstein and J. A. Naglieri (New York, NY: Springer), 301–331. doi: 10.1007/978-1-4614-8106-5_18
- Sorvo, R., Koponen, T., Viholainen, H., Aro, T., Räikkönen, E., Peura, P., et al. (2017). Math anxiety and its relationship with basic arithmetic skills among primary school children. *Br. J. Educ. Psychol.* 87, 309–327. doi: 10.1111/bjep.12151
- Stern, E. (1993). What makes certain arithmetic word problems involving the comparison of sets so difficult for children? *J. Educ. Psychol.* 85, 7–23. doi: 10.1037/0022-0663.85.1.7
- Swanson, H. L., Jerman, O., and Zheng, X. (2008). Growth in working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. *J. Educ. Psychol.* 100, 343–379. doi: 10.1037/0022-0663.100.2.343
- Toplak, M. E., Bucciarelli, S. M., Jain, U., and Tannock, R. (2008). Executive functions: performance-based measures and the behavior rating inventory of executive function (BRIEF) in adolescents with attention deficit/hyperactivity disorder (ADHD). *Child Neuropsychol.* 15, 53–72. doi: 10.1080/09297040802070929
- Toplak, M. E., West, R. F., and Stanovich, K. E. (2013). Practitioner review: do performance-based measures and ratings of executive function assess the same construct? *J. Child Psychol. Psychiatry* 54, 131–143. doi: 10.1111/jcpp.12001
- Tsui, J. M., and Mazzocco, M. M. (2006). Effects of math anxiety and perfectionism on timed versus untimed math testing in mathematically gifted sixth graders. *Roeper Rev.* 29, 132–139. doi: 10.1080/02783190709554397
- Willis, G. B., and Fuson, K. C. (1988). Teaching children to use schematic drawings to solve addition and subtraction word problems. *J. Educ. Psychol.* 80, 192–201. doi: 10.1037/0022-0663.80.2.192
- Wong, B. Y. L. (1986). “Problems and issues in definition of learning disabilities,” in *Psychological and Educational Perspectives on Learning Disabilities*, eds J. K. Torgesen and B. Y. L. Wong (New York: Academic Press), 3–26.
- Wu, S., Amin, H., Barth, M., Malcarne, V., and Menon, V. (2012). Math anxiety in second and third graders and its relation to mathematics achievement. *Front. Psychol.* 3:162. doi: 10.3389/fpsyg.2012.00162
- Zbornik, J. J., and Wallbrown, F. H. (1991). The development and validation of a scale to measure reading anxiety. *Read. Improv.* 28, 2–13.
- Zhang, J., McBride-Chang, C., Wong, A. M.-Y., Tardif, T., Shu, H., and Zhang, Y. (2014). Longitudinal correlates of reading comprehension difficulties in Chinese children. *Read. Writ.* 27, 481–501. doi: 10.1002/dys.1566
- Zhang, J., Meng, Y., Wu, C., and Zhou, D. Q. (2017). Writing system modulates the association between sensitivity to acoustic cues in music and reading ability: evidence from Chinese-English bilingual children. *Front. Psychol.* 8:1965. doi: 10.3389/fpsyg.2017.01965

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

Copyright © 2018 Zhang, Cheung, Wu and Meng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Math Anxiety in Combination With Low Visuospatial Memory Impairs Math Learning in Children

Mojtaba Soltanlou^{1,2,3*}, Christina Artemenko^{1,2}, Thomas Dresler^{2,4},
Andreas J. Fallgatter^{2,4,5}, Ann-Christine Ehlis^{2,4} and Hans-Christoph Nuerk^{1,2,3}

¹ Department of Psychology, University of Tuebingen, Tübingen, Germany, ² LEAD Graduate School & Research Network, University of Tuebingen, Tübingen, Germany, ³ Leibniz-Institut für Wissensmedien, Tübingen, Germany, ⁴ Department of Psychiatry and Psychotherapy, University Hospital of Tuebingen, Tübingen, Germany, ⁵ Center for Integrative Neuroscience, Excellence Cluster, University of Tuebingen, Tübingen, Germany

OPEN ACCESS

Edited by:

Natasha Kirkham,
Birkbeck, University of London,
United Kingdom

Reviewed by:

Annemie Desoete,
Ghent University, Belgium
Lu Wang,
Ball State University, United States

*Correspondence:

Mojtaba Soltanlou
mojtaba.soltanlou@uni-tuebingen.de

Specialty section:

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

Received: 10 March 2018

Accepted: 11 January 2019

Published: 31 January 2019

Citation:

Soltanlou M, Artemenko C,
Dresler T, Fallgatter AJ, Ehlis A-C and
Nuerk H-C (2019) Math Anxiety
in Combination With Low Visuospatial
Memory Impairs Math Learning
in Children. *Front. Psychol.* 10:89.
doi: 10.3389/fpsyg.2019.00089

Math anxiety impairs academic achievements in mathematics. According to the processing efficiency theory (PET), the adverse effect is the result of reduced processing capacity in working memory (WM). However, this relationship has been examined mostly with correlational designs. Therefore, using an intervention paradigm, we examined the effects of math anxiety on math learning. Twenty-five 5th graders underwent seven training sessions of multiplication over the course of 2 weeks. Children were faster and made fewer errors in solving trained problems than untrained problems after learning. By testing the relationship between math anxiety, WM, and math learning, we found that if children have little or no math anxiety, enough WM resources are left for math learning, so learning is not impeded. If they have high math anxiety and high visuospatial WM, some WM resources are needed to deal with math anxiety but learning is still supported. However, if they have high math anxiety and low visuospatial WM capacity, math learning is significantly impaired. These children have less capacity to learn new math content as cognitive resources are diverted to deal with their math anxiety. We conclude that math anxiety not only hinders children's performance in the present but potentially has long-lasting consequences, because it impairs not only math performance but also math learning. This intervention study partially supports the PET because only the combination of high math anxiety and low WM capacity seems critical for hindering math learning. Moreover, an adverse effect of math anxiety was observed on performance effectiveness (response accuracy) but not processing efficiency (response time).

Keywords: math anxiety, math learning, children, individual differences, visuospatial working memory, processing efficiency theory

INTRODUCTION

Math acquisition is influenced by emotional factors such as math anxiety (Dowker et al., 2016). Individuals suffering from math anxiety experience a negative feeling whenever they are presented with mathematics, which impairs their math performance (Devine et al., 2012; Suarez-Pellicioni et al., 2016). Highly math-anxious individuals take a longer time to respond and/or make more errors than individuals with less math anxiety during math problem solving. Supporting the

behavioral findings, neuroimaging studies have shown that math anxiety triggers the fear and hyper-sensitive brain network (for a review see Artemenko et al., 2015). This negative relation between math anxiety and math performance has been explained in different ways. Ashcraft (2002) suggests that highly math-anxious individuals tend to avoid activities and situations that require math. As a consequence, they have less practice with math, which hinders their math knowledge and ability. Another explanation is that highly math-anxious individuals who think that they are bad at math, can be easily distracted during the task (Eysenck et al., 2007) because they do not feel self-confident, and do not allocate their maximum effort to the task (Dowker et al., 2016).

In addition to emotional factors, cognitive processes such as working memory (WM) have been frequently shown to be core determinants for successful learning in school (e.g., Aronen et al., 2005; Lee and Bull, 2016). Lee and Bull (2016) argued that WM is needed while learning new skills including math and also to integrate the new information with previously acquired knowledge. According to Baddeley's model (Baddeley, 1992), WM contains three components: (i) the visuospatial WM, known as the visuospatial sketchpad, which is a transient storage space for visual and spatial information; (ii) the verbal WM, known as the phonological loop, or the transient storage of verbal information; and (iii) the central executive, which is involved in regulating, manipulating, and generally processing the stored information. Prior studies have shown that different WM components play distinct roles in academic achievement during development. For instance, visuospatial WM was a strong predictor of math performance in 7- to 9-year-old children, whereas verbal WM and central executive were not (Holmes and Adams, 2006). Soltanlou et al. (2015) revealed that verbal WM was the best predictor of multiplication performance in grade 3 (8–11 years old); however, visuospatial WM was the best predictor of multiplication performance a year later in grade 4. In general, there is agreement that WM has an integral role in math performance (Menon, 2016; but see Nemati et al., 2017).

Working memory processes *per se* are also influenced by emotional factors such as math anxiety. The literature shows that math anxiety interferes with different WM components. For instance, Passolunghi et al. (2016) observed that children with low math anxiety show a better verbal WM than highly math-anxious children in grades 6 to 8 (11–15 years old). DeCaro et al. (2010) investigated the performance of adults on two kinds of verbal WM-based and visual WM-based math tasks during low- and high-pressure testing situations. The authors found that while a high-pressure situation attenuated the performance in a verbal WM-based math task, it was not influential in the visual WM-based task. They suggested that anxiety has a greater influence on verbal WM rather than visual WM. However, several other studies suggest a selective disruption effect of anxiety on visual WM in adults (Miller and Bichsel, 2004; Shackman et al., 2006) and in children in grades 1 and 2 (7–9 years old) (Vukovic et al., 2013). Despite these inconsistent findings across the literature, there is general agreement that anxious thoughts partially occupy WM capacities, which disrupts math performance.

As mentioned above, math anxiety, WM, and math performance are related to each other, whereby WM has been suggested to mediate the anxiety-performance relationship (cf. Figure 1). The processing efficiency theory (PET, Eysenck and Calvo, 1992) offers a good explanation for the interaction between them. The PET was developed based on Baddeley's model of WM (Baddeley, 1992) and suggests that anxiety causes worry, which reduces the WM capacity, disrupting concurrent tasks. It contains two main concepts: performance effectiveness and processing efficiency (Eysenck and Calvo, 1992). Performance effectiveness refers to the quality of performance, i.e., the response accuracy, while processing efficiency refers to the relationship between performance effectiveness and a load of effort or cognitive resources, i.e., response time. For instance, occupying WM capacity leads to performance impairment (affecting performance effectiveness), but availability of auxiliary cognitive resources maintains a given performance level but at the cost of increased effort (affecting processing efficiency). Therefore, according to the PET, WM might be the best intermediate variable explaining the relationship between math anxiety and math performance.

There are two different accounts regarding the interaction of math anxiety, WM and math performance. One account is that individuals with higher WM capacity have more resources to simultaneously manage math anxiety and solve math problems (Ashcraft and Kirk, 2001). For example, a study in 11- to 12-year-old children reported that verbal WM accounts for 51% of the association between trait anxiety and academic performance including math (Owens et al., 2008). Therefore, children with low WM capacity suffer more from math anxiety during math problem solving. The other account suggests that individuals with higher WM capacity suffer more from math anxiety (Beilock and Carr, 2005) because they rely heavily on WM strategies to solve math problems. Therefore, under any high-pressure situation, their capacity is co-opted and they show a worse performance (Ramirez et al., 2013). This deficit does not occur for individuals with lower WM capacity because they do not rely massively on WM strategies to solve math problems in the first place, but rather use other strategies. Therefore, their performance does not drastically diminish in high-pressure situations. For instance, Ramirez et al. (2013) reported a relationship between math anxiety and verbal WM in children with higher WM capacity in grades 1 and 2 (see also Vukovic et al., 2013). So, despite contradictory findings across mediation studies, they mostly agree on the mediating role of WM in the association between math anxiety and math performance.

Although these relationships have been frequently studied, most of our knowledge comes from correlational studies, which have investigated the influence of math anxiety on a single measure of math performance. Therefore, longitudinal (e.g., Vukovic et al., 2013; Cargnelutti et al., 2016) and intervention studies are needed to clarify the causality of these relationships (Dowker et al., 2016). While correlational studies reveal possible associations between two variables, causal studies indicate the directionality of these associations. For instance, correlational studies revealed that math anxiety is associated with poor performance in both WM and math tasks. However, this

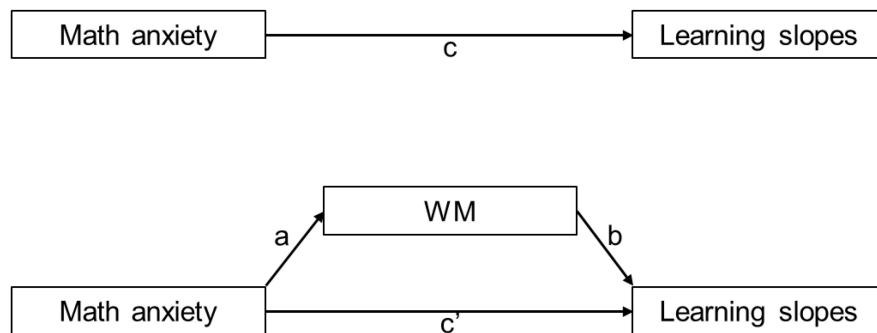


FIGURE 1 | Path diagram: above panel depicts the total effect of the predictor (math anxiety) on the dependent variable (learning slopes) and below panel depicts the direct effect of the predictor on dependent variable while controlling for the mediator (WM components) and the indirect effect of the predictor on the dependent variable through the mediator.

relationship can be bidirectional: (i) math anxiety preoccupies WM and individuals attend less to the task (Eysenck et al., 2007), which leads to a low score on WM and math tasks (Ashcraft and Kirk, 2001), and (ii) poor math knowledge makes individuals worry because they feel incapable of solving math problems, so they show a high score on math anxiety tests (Maloney et al., 2011; Núñez-Peña and Suárez-Pellicioni, 2014; Lindskog et al., 2017). Therefore, the perennial “chicken and egg” question will not be resolved by correlational studies and intervention studies are needed (Dowker et al., 2016).

In one of the few longitudinal studies, Cargnelutti et al. (2016) observed that math anxiety and math performance have a bidirectional relationship. Nevertheless, math performance has a greater impact on math anxiety in 2nd graders (7–9 years old), whereas the reverse directionality was observed a year later in 3rd graders. Interestingly, they observed an indirect effect of math anxiety in 2nd graders on math performance in 3rd graders, suggesting poor math skills may cause math anxiety in younger children that disrupts math performance later. Supporting this finding, Ma and Xu (2004) suggested that prior math achievement longitudinally predicts later attitudes toward math across grades 7 to 12. However, the influence of WM on the association between math anxiety and performance was not investigated in these studies. Another longitudinal study (Vukovic et al., 2013) investigated this relationship by taking into account the WM capacity. The authors observed that high math anxiety in 2nd graders predicts less math acquisition from grade 2 to grade 3 but only in children with higher visuospatial WM capacity. Vukovic et al. (2013) suggested that math anxiety causes poor math learning by affecting WM resources in school children.

Longitudinal studies, however, also come with the possible confounding effects of brain maturation and concurrent economic trends or other events affecting children’s lives over a long timescale. Therefore, the findings of training studies might differ from longitudinal studies (Soltanlou et al., 2018). Accordingly, we conducted an intervention study in children to uncover the association between math anxiety and math learning, namely the difference in competence before and after learning. Furthermore, the possible mediating roles of different WM components were tested. We hypothesized that higher math

anxiety leads to less benefit from arithmetic learning, and that this relationship is modulated by WM.

MATERIALS AND METHODS

Participants

Twenty six typically developing children from 5th grade participated in the study. One child, who quitted training, was excluded and the remaining 25 children (9 girls; 11.13 ± 0.46 years old) were included in the analyses. All children were right-handed and had normal or corrected-to-normal vision with no history of neurological or mental disorders. Intellectual ability was measured by completing two subtests (similarities and matrix reasoning) of the German version of the Wechsler Intelligence Scale (Petermann et al., 2007), with resulting scores of 107.40 ± 11.65 and 107.80 ± 10.61 , respectively. Children and their parents gave written informed consent and received an expense allowance for their participation. All procedures of the study were in line with the latest revision of the Declaration of Helsinki and were approved by the ethics committee of the University Hospital of Tuebingen.

Material

Math Anxiety

Math anxiety was assessed by selected items from the German translation of the math anxiety questionnaire (MAQ) (Thomas and Dowker, 2000; Krinzinger et al., 2007), which has an internal consistency (Cronbach’s alpha) of 0.83–0.91 for the whole questionnaire for different age groups. In the questionnaire, we assessed three out of four subscales of the MAQ: self-assessment in math, attitude toward math, and concerns about math¹. In our questionnaire, each subscale

¹Krinzinger and colleagues (one of which – Nuerk – is co-author both of her and our paper), developed the German version of the MAQ, which we used in our study. Their results strongly suggested that the two subscales of “How happy or unhappy are you if you have problems with...?” and “How worried are you if you have problems with...?” actually measure the same construct, which is negative emotions and anxiety concerning mathematics (see also Krinzinger et al., 2009). Krinzinger et al. (2007) observed a correlation of 0.78 between these two

contains five items describing different math-related topics (calculation, handwritten calculation, mental calculation, simple calculation problems, and difficult calculation problems). While the subscales self-assessment in math and attitude toward math demonstrate general math-related attitudes, the subscale concerns about math indicates math anxiety (Krinzinger et al., 2009). Since we are only interested in the influence of math anxiety on math learning, we focus on the last subscale hereafter. This subscale includes five items, which are rated on a five-point Likert scale (ranging from 0 = very happy to 4 = very unhappy) with a maximum score of 20. Thereby, higher values indicate higher math anxiety.

Working Memory

Following Baddeley's model (Baddeley, 1992), three components of WM, i.e., verbal WM, visuospatial WM, and central executive were measured. To this end, the letter span test (Soltanlou et al., 2015) and the Corsi block-tapping test (Corsi, 1973) were used. In the letter span test, the child had to recall spoken sequences of letters (presentation rate: one letter per second). The test was started with sequences of two letters. The sequence length was increased by one letter if the child recalled correctly at least one out of two sequences; otherwise, testing was stopped. In the Corsi block-tapping test, the child was asked to point to the cubes in the same order as the experimenter. Children started with sequences of three cubes. The sequence length was increased by one cube if the child recalled correctly at least two out of three sequences; otherwise, testing was stopped. For the backward in both tasks, children were asked to recall sequences in reverse order. The forward and backward spans are distinguishable and related differentially to math performance in children (Soltanlou et al., 2015).

Hoshi et al. (2000) revealed that backward span leads to greater activation in the bilateral prefrontal cortex than forward span. Therefore, the forward span in the letter span test represents the verbal WM, and the forward span in the Corsi block-tapping test represents visuospatial WM. For both forward and backward span of both verbal and visuospatial WMs, the score was the maximum sequence length at which at least two sequences were repeated correctly. The average of the backward spans of the two tests represents the central executive. Note that the backward span of the letter span test (e.g., Hadwin et al., 2005) and the backward span of the Corsi block-tapping test (e.g., Vandierendonck et al., 2004) have been separately reported as measures of the central executive. Vandierendonck et al. (2004) state a similar involvement of the central executive in the backward span of the Corsi block-tapping test and the backward letter/digit span (Vandierendonck et al., 1998). Moreover, according to the theoretical definition, the central executive is modality-independent (Baddeley, 1992) and is involved in manipulating both verbal and visual information.

constructs in third graders, which is close to reliability. Since this means that we essentially measure the same construct twice, we only used the first subscale in the present study. As this study is a part of a larger project of math training in children, we had to shorten some tests by only concentrating on the most relevant parts. Therefore, these two MAQ subscales, which essentially measured the same construct in German children, were the natural candidates for that.

Therefore, the average of the backward span in the letter span and the Corsi block-tapping tests, which are functionally similar (Logie, 2014), was considered to be an indicator of the central executive in the current study. The internal consistency (Cronbach's alpha) is 0.79 and 0.70–0.79 for the letter span (Kane et al., 2004) and the Corsi block-tapping test (Orsini, 1994), respectively.

Multiplication

In the present study, 16 simple and complex multiplication problems were used. Half of the problems of each set were used as trained problems and the other closely matched half were used as untrained problems. The sets were matched based on the sizes of the operands and results, as well as the parity of the operands and results, separately for simple and complex multiplication problems. The simple problems (e.g., 3×7) included two single-digit operands (range 2–9) with two-digit solutions (range 12–40). The complex problems included one two-digit operand (range 12–19) and one single-digit operand (range 3–8) with a two-digit solution (range 52–98). The sequence of small and large operands within the problems was counterbalanced. Problems with ones (e.g., 9×1), commutative pairs (e.g., 3×4 and 4×3) or ties (e.g., 6×6) were not used (for more see Soltanlou et al., 2018). According to the PET, which suggests the effect of math anxiety on complex tasks, and because of our small sample size, we only report the findings of complex multiplication problems. Trained and untrained multiplication task in the pre-training and post-training sessions has an internal consistency (Cronbach's alpha) of 0.82 in the current study.

Procedure Measurement

This study is a part of a larger behavioral and neuroimaging project on math learning in children (Soltanlou et al., 2017, 2018). In a within-subject experiment, math performance of children was measured before and after training in both trained and untrained complex multiplication problems. The IQ, MAQ, and WM measures were administered after the post-training measure. Measurement of math anxiety after the math task has the advantage of avoiding any possible pre-judgment and bias about the forthcoming task in children (see also Ramirez et al., 2013). The math task was preceded by four practice trials. Problems were presented on a touch screen and children had to write their answers as quickly and accurately as possible and then in order to continue, they needed to click on a gray box presented on the right side of the screen (see Soltanlou et al., 2018 for more details). The written response was not visible to avoid any further corrections and to encourage children to calculate mentally. The problems of each set were presented in four blocks of 45 s, each followed by 20 s of rest. The sequence of blocks and problems within the blocks was pseudo-randomized. The problems, but not the sequence of the blocks or problems, were identical for each set in pre-training and post-training sessions. Whenever the total number of trials within a set was reached, the same problems were presented again after randomization. No feedback was given during the experiment. The design was self-paced with a limited response interval of 30 s

for each problem. Therefore, due to inter-individual differences, the number of solved problems varied between children. The inter-trial interval was set to 0.5 s. The experiment was run using Presentation® software version 16.3 (Neurobehavioral Systems Inc.).

Training

Training was conducted via an online learning platform (Jung et al., 2015, 2016; Roesch et al., 2016), which allows for at-home training. The problems in the trained complex multiplication condition were randomly repeated six times in each training session. Each problem was individually presented along with 12 different choices including the correct solution (see Soltanlou et al., 2018). Response intervals of complex problems ranged randomly between 10 and 30 s, jittered by 2 s. Whenever the child did not respond within the response interval, the computer screen displayed the correct solution. Training was interactive because children had to compete with the computer. In order to create a more realistic competition, the computer responded incorrectly in 30% of the problems. To provide immediate feedback about the performance and to increase motivation, the scores of the child and computer were shown on the right side of the screen after choosing a solution. Both child and computer received one point for each correct answer and one point was deducted for each incorrect answer. The problem was presented until the child or computer responded correctly. Children were instructed to solve the problems as quickly and accurately as possible. Children performed seven sessions of approximately 25-min interactive training between two measurement times: one session in the lab and six sessions at home during about 2 weeks. The post-training session was conducted after these 2 weeks.

Analysis

For the math task, the written responses by children were read out with the help of the RON program (Ploner, 2014). Response times (RTs) were defined as the time from problem presentation to pressing the gray box. Only mean RTs for correct responses (74.45% of problems across both measurement times) were included in the analyses. Error rate was defined as the proportion of incorrect or missing responses to the total number of presented trials. Furthermore, in order to approximate a normal distribution, an arcsine-square-root-transformation of error rate (Winer et al., 1971) was calculated.

Thereafter, learning slopes were calculated by subtracting the mean RT and arcsine-square-root-transformed error rates of the pre-training session by post-training session separately for trained and untrained multiplication sets for each child. In both RT and error rate, larger values show higher training effects. Paired *t*-tests were conducted between trained and untrained sets for both RT and error rate learning slopes separately.

In order to test the associations between variables, correlation and regression analyses were calculated. Based on these analyses, mediation analysis was conducted by considering math anxiety as a predictor, learning slopes as dependent variables, and any WM component that significantly correlated with math anxiety, as a mediator (cf. **Figure 1**). According to Baron and Kenny's (1986) causal-steps test (1986), four assumptions need to be met for mediation analysis (see also Field, 2013): (1) the total effect of a predictor on the dependent variable (path c) must be significant, (2) the effect of predictor on mediator (path a) must be significant, (3) the effect of mediator on dependent variable (path b), while controlled for predictor, must be significant, (4) the direct effect of predictor on dependent variable (path c'), while controlled for mediator, must be smaller than the total effect of predictor on dependent variable (path c) (cf. **Figure 1**). However, more liberal mediation tests such as the joint significance test (MacKinnon et al., 2002) suggest that only the second and third assumptions are required and the first and fourth assumptions are not necessary (for more see Fritz and MacKinnon, 2007). The Sobel test or delta method was used for the mediation analysis. This method estimates the standard error of the indirect effect and assumes the sampling distribution of the indirect effect as being normal². It assesses the presence of mediation by dividing the indirect effect by the first-order delta-method standard error of the indirect effect and then compares it against a standard normal distribution. If the result of this calculation is significant, mediation is present (Fritz and MacKinnon, 2007). The analysis was completed using RStudio (RStudio Team, 2016) and jamovi software (jamovi project, 2018).

²Note that the non-parametric percentile bootstrap confidence interval method with 5000 samples, which does not assume a normal distribution of the indirect effect, revealed a similar suppression effect in mediation analysis (see the Result section).

TABLE 1 | Correlation between math anxiety, WM components, and learning slopes.

	Mean (SD)	Range	Correlations					
			1	2	3	4	5	6
1. Math anxiety	12.92 (3.17)	0–20	–	–0.12	–0.43*	0.02	–0.08	–0.31
2. Verbal WM	4.92 (1.04)	2–9		–	–0.14	0.35*	–0.10	0.15
3. Visuospatial WM	5.00 (0.58)	3–9			–	0.22	0.00	–0.20
4. Central executive	4.52 (0.81)	2.5–8.5				–	–0.12	0.17
5. RT learning slope	4.27 (3.06)	–					–	0.66**
6. Error rate learning slope	0.11 (0.20)	–						–

N = 25, ***p* < 0.01, **p* < .05, two-tailed.

RESULTS

Learning Slopes

A paired *t*-test on the RT learning slopes revealed a significant training effect in trained problems ($M = 4.27$ s, $SD = 3.06$ s) compared to untrained problems ($M = 1.60$ s, $SD = 2.41$ s), $t(24) = 3.91$, $p < 0.001$, showing that children responded faster to the trained set than untrained set due to training. A paired *t*-test on the error rate learning slopes again revealed a significant training effect in trained problems ($M = 0.11$, $SD = 0.20$) compared to untrained problems ($M = -0.04$, $SD = 0.18$), $t(24) = 3.30$, $p = 0.003$, showing that children made less errors when solving trained problems than untrained problems due to training.

Correlation and Regression

The correlation and regression analyses revealed the following results. (1) No significant correlations between math anxiety and learning slopes (path c) were observed. (2) A negative correlation between math anxiety and visuospatial WM (path a) showed higher anxiety with decreasing visuospatial WM. Since math anxiety only correlated with visuospatial WM, further analyses were conducted only on this WM component.

Additionally, significant correlations between verbal WM and central executive, and between RT learning slope and error rate learning slope, were observed. No other significant correlations were observed (cf. **Table 1**).

(3) Regression analysis to test the effect of visuospatial WM on error rate learning slope while controlling for math anxiety (path b) was only marginally significant, $R^2 = 0.23$, $F(2,22) = 3.30$, $p = 0.056$ (cf. **Table 2**). The result revealed that the higher math anxiety and the higher visuospatial WM (but marginally significant) the lower math learning as indicated by error rates. This finding shows a suppression effect: while neither math anxiety nor visuospatial WM correlated with error rate learning slope, by inserting them together, they significantly predicted error rate learning slope. A suppression effect is defined when adding the third variable (i.e., WM) increases the effect of the independent variable (i.e., math anxiety) on the dependent variable (i.e., learning), which is the opposite effect of the third variable in mediation.

TABLE 2 | Regression analyses (path b).

	<i>b</i>	<i>SE</i>	<i>B</i>	<i>t</i>	<i>p</i>
Math anxiety and visuospatial WM → Error rate learning slope					
Constant	1.19	0.45	0.00	2.64	0.015*
Math anxiety	−0.03	0.01	−0.48	−2.33	0.030*
Visuospatial WM	−0.14	0.07	−0.41	−1.98	0.060
Math anxiety and visuospatial WM → RT learning slope					
Constant	6888.40	7977.60	0.00	0.86	0.397
Math anxiety	−101.70	226.80	−0.11	−0.45	0.658
Visuospatial WM	−206.70	1246.90	−0.05	−0.21	0.836

$N = 25$; *b*, unstandardized beta coefficient; *SE*, standard error of *b*; *B*, standardized beta coefficient. * $p < 0.05$.

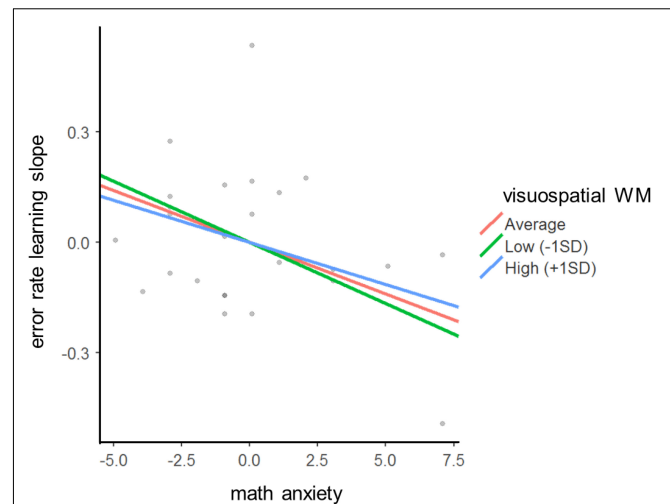


FIGURE 2 | Simple slope analysis showing the effect of math anxiety on error rate learning slope at three different levels of visuospatial WM capacity. The scores of math anxiety and error rate learning slopes have been centered by subtracting each value from the mean. While children with low and average visuospatial WM capacity were significantly influenced by math anxiety and drew less benefit from multiplication learning, children with high visuospatial WM capacity were not significantly influenced by math anxiety.

Regression analysis to test the effect of visuospatial WM on RT learning slope while controlling for math anxiety (path b) was not significant, $R^2 = 0.01$, $F(2,22) = 0.10$, $p = 0.905$ (cf. **Table 2**). Since this assumption was not met for RT learning slope, further analysis was conducted only on error rate learning slope.

(4) The mediation analysis revealed that by inserting visuospatial WM as the mediator to the model, math anxiety significantly predicts (path c') error rate learning slope (cf. **Table 3**). The suppression effect was also corroborated by this finding that the estimation of the total effect (path c) is closer to zero than the direct effect (path c'), and the estimation of direct and indirect effects have opposite signs (MacKinnon et al., 2000).

In order to explore the relationship between these three variables, a simple slopes analysis (Aiken and West, 1991) was conducted on the z-transformed scores. According to the simple slopes analysis, the effect of math anxiety on error rate learning slope is investigated at low, average, and high levels of visuospatial WM capacity. As a standard method, low and high levels are defined as 1 SD below and above the mean, respectively. The analysis revealed that children with low ($b = -0.03$, $z = -2.60$, $p = 0.009$) and average ($b = -0.03$, $z = -2.38$, $p = 0.017$) visuospatial WM capacity were significantly influenced by math anxiety and got less benefit from multiplication learning (cf. **Figure 2**), while children with high visuospatial WM capacity are not significantly influenced by math anxiety ($b = -0.02$, $z = -1.19$, $p = 0.233$).

DISCUSSION

In the present intervention study, children improved after seven sessions of complex multiplication training. Moreover, an

TABLE 3 | Mediation analysis between math anxiety, visuospatial WM, and error rate learning slope.

Effect	Label	<i>b</i>	SE	95% Confidence interval		<i>Z</i>	<i>p</i>
				Lower	Upper		
Indirect	$a \times b$	0.011	0.007	−0.003	0.024	1.58	0.113
Direct	c'	−0.030	0.012	−0.053	−0.006	−2.48	0.013*
Total	c	−0.019	0.011	−0.042	0.004	−1.60	0.109

N = 25; *b*, unstandardized beta coefficient; SE, standard error of *b*. **p* < 0.05.

association between math anxiety, visuospatial WM, and math learning was observed.

We observed a significant negative relationship between math anxiety and visuospatial WM, suggesting that children with higher math anxiety have less storage capacity for visual and spatial information. This finding is in line with previous literature reporting the influence of math anxiety on visuospatial WM (e.g., Trezise and Reeve, 2018). Miller and Bichsel (2004) found math anxiety effects on visual WM but not on verbal WM. They suggest that while other types of anxiety affect verbal processes, math anxiety has a different unique effect on visual WM. In a similar way, Shackman et al. (2006) observed that anxiety selectively disrupts visuospatial WM but not verbal WM. However, the adverse effect of anxiety on other WM components has been shown as well. For instance, Hadwin et al. (2005) observed that low-anxious children aged 9–10 years old were faster in doing forward and backward digit span tasks (verbal WM and central executive) than high-anxious children, but not in a visuospatial WM task. Our finding suggests that because 5th graders rely on their visuospatial WM to solve multiplication problems, if math anxiety has any effect, this effect might be on this skill rather than verbal WM.

Although literature reported a strong association between WM and math performance (Aronen et al., 2005; Menon, 2016), we did not observe this relationship in the correlation analysis. However, visuospatial WM was a nearly significant predictor of error rate learning slope when we added math anxiety to the model. This finding might point to the necessity of math anxiety as an individual difference measure, which needs to be taken into account when we investigate math acquisition during development (Vukovic et al., 2013). As Vukovic et al. (2013) suggest, math anxiety influences how children utilize their WM capacity to learn math. The importance of visuospatial WM in multiplication problem solving has already been shown in children (Soltanlou et al., 2015, 2017). Unexpectedly, the relationship between visuospatial WM and error rate learning slope was negative, showing that children with higher visuospatial WM get less benefit out of multiplication learning. One interpretation might be because they had already few errors in pre-training, therefore, this short training did not lead to a significant improvement in these children. However, this association will be disambiguated later by exploring the interaction between math anxiety, visuospatial WM, and error rate learning slope.

Interestingly, by adding both math anxiety and visuospatial WM as predictors of math learning, a suppression effect was

observed: the influence of math anxiety on math learning increased by adding visuospatial WM to the regression model. When exploring this relationship, we observed that while children with a low and average capacity of visuospatial WM are more influenced by math anxiety, children with a high visuospatial WM capacity can compensate the negative influence of math anxiety on learning. As Ashcraft and Kirk (2001) suggested, individuals with higher WM capacity have more resources to simultaneously deal with math anxiety and solve the math problems (see also Miller and Bichsel, 2004). The general pattern of findings – from the simple slope analysis – is partially in line with the study by Owens et al. (2012). They showed that trait anxiety is negatively correlated with cognitive performance in 12- to 14-year-old children with low WM capacity; however, no significant correlation was observed in children with average WM capacity. Contradictory to our findings, they found a positive relationship between trait anxiety and cognitive performance in children with high WM capacity.

It seems that the combination of high math anxiety and low WM is critical for hindering math learning. One might argue that children with high WM capacity have enough resources to attenuate the influence of math anxiety on math acquisition, which is in line with the PET. We suggest that this claim is correct if WM mediates the association between math anxiety and math learning, similar to several correlational studies. These studies revealed that either verbal WM (e.g., Owens et al., 2008) or visuospatial WM (e.g., Miller and Bichsel, 2004) mediates the anxiety-math performance association. There is a crucial conceptual difference between mediation and suppression: while WM reduces the influence of math anxiety on math performance in mediation, this effect increases in suppression³. So, while the correlational studies found the former, we observed the latter in our learning study. Furthermore, as Hopko et al. (2003) discussed, a single measure of math performance at a certain time is not purely a measure of competence, but a measure of both math anxiety and competence combined. Individuals start solving math problems with different levels of math anxiety, which is most probably represented in their output as well. We conclude that the findings of correlational studies may not be readily generalized to causal and intervention studies.

Furthermore, we found that math anxiety had a negative influence on children with low and average WM capacity but this

³Note that this relationship was not a moderation because (1) in moderation the relationship between predictor and dependent variable is significant *per se* but it changes by the third variable, however, this relationship was not significant in our data, (2) the interaction of math anxiety and visuospatial WM (moderation analysis) did not significantly predict learning slopes.

influence was not significant in children with high WM capacity. As we explained in the introduction, there are two contradictory accounts of the relationship between math anxiety and WM capacity across the literature: one suggests that math anxiety has a negative impact in individuals with low WM capacity (Ashcraft and Kirk, 2001); the other suggests that individuals with higher WM capacity suffer more from math anxiety (Beilock and Carr, 2005). Our findings adhere to the first account, showing that children with higher WM capacity have enough resources to deal simultaneously with anxious thoughts and also store and manipulate new information (Eysenck et al., 2007). As Lee and Bull (2016) argued, WM is needed when learning new academic skills to integrate the new information with previously acquired knowledge. This explanation is corroborated by neuroimaging studies revealing increased prefrontal activation for emotion regulation, in addition to the fundamental role of the right amygdala in emotion processing (Young et al., 2012). Therefore, prefrontal capacity that subserves cognitive processes such as WM is partially allocated to regulate these affective responses. Hence, this capacity is less available for the cognitive task at hand, such as solving a math problem (Eysenck and Calvo, 1992; Eysenck et al., 2007). Therefore, it is reasonable to see a stronger association between math anxiety and math learning in children with lower WM capacity.

Inconsistent with the PET, performance effectiveness (response accuracy) and not processing efficiency (response time) was influenced by math anxiety in our intervention study. The prediction of the PET has received support and contradictory evidence in the field of numerical cognition. For instance, Ng and Lee (2010) observed that processing efficiency – but not performance effectiveness – on a mental arithmetic task is affected by test anxiety in 10-year-old children. Vukovic et al. (2013), however, observed a negative correlation between math anxiety and performance effectiveness in their longitudinal study, which supports our findings (see also Devine et al., 2012). Nonetheless, they did not measure the response time in their math tasks, which might have shown a significant association as well. In line with their finding, Trezise and Reeve (2018) showed that while anxiety is negatively related to the response accuracy in two low- and high-time pressure conditions, there is no significant correlation between math anxiety and response time in 14-year-old children. It seems that the underlying mechanisms of one-time math performance measures differ from math learning. We suggest that – in line with the PET – a negative correlation between math anxiety and math learning was observed in the present study; however, contradictory to its prediction, this relationship was between anxiety and response accuracy, and not response time.

Limitations

There are some limitations that need to be taken into account for interpretation of our findings which should be addressed in future studies. Our study was a complex and effortful intervention study, in which not so many children can be easily tested, as compared to cross-sectional correlational designs. Therefore, null effects in particular were and should

be interpreted with caution due to low power. Especially, if there are smaller intervention or mediation effect sizes, it is conceivable that they might be observed in a larger sample. Moreover, in order to reduce confounding effect of maturation and education, we conducted this study in a group of 5th graders with a limited age range. Therefore, the influence of math anxiety on learning, which we observed here, needs to be further investigated in larger samples and in different age groups to see whether our findings can be replicated and generalized.

Moreover, it is suggested to measure the other types of anxieties to see whether our findings are math specific or related to trait or test anxiety as well. Although we investigated several other interesting factors such as gender, task complexity, and self-attitude in our study, however, because of the small sample size, we focus only on the most important question: whether math anxiety influence on math learning in children. Therefore, it is suggested for future studies to consider these factors as well.

CONCLUSION

Most studies so far have only investigated the influence of math anxiety and WM on math performance. In such studies, both variables have a negative impact on math performance, and in some studies (in line with the PET) WM mediates the influence of math anxiety on math performance.

Our study suggests that the case might be different for the influence of math anxiety and WM on *math learning*. While an influence of WM on math performance is ubiquitous, we failed to find a significant influence of any of the WM components on math learning. This might be partially consistent with a recent meta-analysis showing that WM training does not transfer strongly to other skills and capabilities like math (Melby-Lervag et al., 2016). So, if a child has a higher WM capacity or even if WM is improved after training, he might have a good math performance – in both pre- and post-training measures – but not necessarily improves dramatically after math learning as compared to pre-training performance.

While WM might not predict math learning *per se*, it fosters the influence of math anxiety on math learning. Children with a low visuospatial WM capacity suffer most from math anxiety when they have to learn math. The explanation for this is in line with the PET. If children have no or little math anxiety, enough WM resources are left for math learning, so no major problems occur. If they have high math anxiety and high visuospatial WM, some WM resources are needed to deal with math anxiety but learning is still supported. However, if they have high math anxiety and low visuospatial WM capacity, math learning is significantly impaired. These children have less capacity to learn new math contents because they need all the resources to deal with their math anxiety. This finding might be helpful for future interventions and suggests that in order to improve children's performance, both math anxiety and WM capacity need to be considered.

Our findings show that math anxiety plays a major role in multiplication learning and that data from performance studies cannot be readily generalized to learning studies. However, multiplication learning is a rather easy task (even if the problems are difficult). The picture might change for other math content. Our study suggests that it is worthwhile to examine the influence of math learning in other math areas as well. After all, learning math is what all children are asked to achieve and where many children suffer tremendously. Therefore, although intervention studies are hard to conduct, we believe it is a worthy and necessary effort to be addressed in future studies if we want to understand and promote math learning in children.

AUTHOR CONTRIBUTIONS

All authors designed and conceptualized the study. MS and CA collected the data. MS analyzed the data and wrote the main manuscript text. All authors reviewed the manuscript.

REFERENCES

- Aiken, L. S., and West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage.
- Aronen, E., Vuontela, V., Steenari, M.-R., Salmi, J., and Carlson, S. (2005). Working memory, psychiatric symptoms, and academic performance at school. *Neurobiol. Learn. Mem.* 83, 33–42. doi: 10.1016/j.nlm.2004.06.010
- Artemenko, C., Daroczy, G., and Nuerk, H.-C. (2015). Neural correlates of math anxiety—an overview and implications. *Front. Psychol.* 6:1333. doi: 10.3389/fpsyg.2015.01333
- Ashcraft, M. H. (2002). Math anxiety: personal, educational, and cognitive consequences. *Curr. Dir. Psychol. Sci.* 11, 181–185. doi: 10.1111/1467-8721.00196
- Ashcraft, M. H., and Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *J. Exp. Psychol. Gen.* 130, 224–237. doi: 10.1037/0096-3445.130.2.224
- Baddeley, A. (1992). Working memory. *Science* 255, 556–559. doi: 10.1126/science.1736359
- Baron, R. M., and Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J. Pers. Soc. Psychol.* 51, 1173–1182. doi: 10.1037/0022-3514.51.6.1173
- Beilock, S. L., and Carr, T. H. (2005). When high-powered people fail: working memory and “choking under pressure” in math. *Psychol. Sci.* 16, 101–105. doi: 10.1111/j.0956-7976.2005.00789.x
- Cargnelutti, E., Tomasello, C., and Passolunghi, M. C. (2016). How is anxiety related to math performance in young students? A longitudinal study of Grade 2 to Grade 3 children. *Cogn. Emot.* 31, 755–764. doi: 10.1080/02699931.2016.1147421
- Corsi, P. M. (1973). *Human Memory and the Medial Temporal Region of the Brain*. Ann Arbor, MI: ProQuest Information & Learning.
- DeCaro, M. S., Rotar, K. E., Kendra, M. S., and Beilock, S. L. (2010). Diagnosing and alleviating the impact of performance pressure on mathematical problem solving. *Q. J. Exp. Psychol.* 63, 1619–1630. doi: 10.1080/17470210903474286
- Devine, A., Fawcett, K., Szűcs, D., and Dowker, A. (2012). Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety. *Behav. Brain Funct.* 8:33. doi: 10.1186/1744-9081-8-33
- Dowker, A., Sarkar, A., and Looi, C. Y. (2016). Mathematics anxiety: what have we learned in 60 years? *Front. Psychol.* 7:508. doi: 10.3389/fpsyg.2016.00508
- Eysenck, M. W., and Calvo, M. G. (1992). Anxiety and performance: the processing efficiency theory. *Cogn. Emot.* 6, 409–434. doi: 10.1080/02699939208409696
- Eysenck, M. W., Derakshan, N., Santos, R., and Calvo, M. G. (2007). Anxiety and cognitive performance: attentional control theory. *Emotion* 7, 336–353. doi: 10.1037/1528-3542.7.2.336
- Field, A. (2013). *Discovering Statistics using IBM SPSS Statistics*. Newcastle upon Tyne: Sage.
- Fritz, M. S., and MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychol. Sci.* 18, 233–239. doi: 10.1111/j.1467-9280.2007.01882.x
- Hadwin, J. A., Brogan, J., and Stevenson, J. (2005). State anxiety and working memory in children: a test of processing efficiency theory. *Educ. Psychol.* 25, 379–393. doi: 10.1080/01443410500041607
- Holmes, J., and Adams, J. W. (2006). Working memory and children’s mathematical skills: implications for mathematical development and mathematics curricula. *Educ. Psychol.* 26, 339–366. doi: 10.1080/01443410500341056
- Hopko, D. R., McNeil, D. W., Lejuez, C., Ashcraft, M. H., Eifert, G. H., and Riel, J. (2003). The effects of anxious responding on mental arithmetic and lexical decision task performance. *J. Anxiety Disord.* 17, 647–665. doi: 10.1016/S0887-6185(02)00240-2
- Hoshi, Y., Oda, I., Wada, Y., Ito, Y., Yamashita, Y., Oda, M., et al. (2000). Visuospatial imagery is a fruitful strategy for the digit span backward task: a study with near-infrared optical tomography. *Cogn. Brain Res.* 9, 339–342. doi: 10.1016/S0926-6410(00)00006-9
- jamovi project (2018). *Jamovi (Version 0.8)*. [Computer Software]. Available at: <https://www.jamovi.org>
- Jung, S., Huber, S., Heller, J., Grust, T., Möller, K., and Nürk, H.-C. (2016). Die tuebingen lernplattform zum erwerb numerischer und orthografischer kompetenzen (TULPE): individualisierte förderung durch adaptive lernspiele. *Lernen Lernstörungen* 5, 7–15. doi: 10.1024/2235-0977/a000112
- Jung, S., Roesch, S., Huber, S., Heller, J., Grust, T., and Nuerk, H. (2015). “An interactive web-based learning platform for arithmetic and orthography,” in *Paper Presented at the Advances in Computers and Technology for Education—Proceedings of the 11th International Conference on Educational Technologies*, eds N. E. Mastorakis, A. L. Brooks, and I. J. Rudas (Daubai: WSEAS), 13–22.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., and Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *J. Exp. Psychol. Gen.* 133, 189–217. doi: 10.1037/0096-3445.133.2.189
- Krinzinger, H., Kaufmann, L., Dowker, A., Thomas, G., Graf, M., Nuerk, H.-C., et al. (2007). Deutschsprachige version des fragebogens für rechenangst (FRA) für 6-bis 9-jährige kinder. *Z. Kinder Jugendpsychiatr. Psychother.* 35, 341–351. doi: 10.1024/1422-4917.35.5.341

FUNDING

This research was funded by a grant from the Science Campus Tuebingen, project 8.4 to H-CN supporting MS. MS was also supported by the DFG grant (NU 265/3-1) to H-CN. All authors are members of the LEAD Graduate School & Research Network (GSC1028), which is funded within the framework of the Excellence Initiative of the German federal and state governments. Furthermore, A-CE was partly supported by the IZKF Tuebingen (Junior Research Group, Grant 2115-0-0).

ACKNOWLEDGMENTS

We would like to thank all participating children and their parents. We also thank our student assistants who helped in data collection and language proofreading of the manuscript.

- Krinzinger, H., Kaufmann, L., and Willmes, K. (2009). Math anxiety and math ability in early primary school years. *J. Psychoeduc. Assess.* 27, 206–225. doi: 10.1177/0734282908330583
- Lee, K., and Bull, R. (2016). Developmental changes in working memory, updating, and math achievement. *J. Educ. Psychol.* 108, 869. doi: 10.1037/edu0000090
- Lindskog, M., Winman, A., and Poom, L. (2017). Individual differences in nonverbal number skills predict math anxiety. *Cognition* 159, 156–162. doi: 10.1016/j.cognition.2016.11.014
- Logie, R. H. (2014). *Visuo-Spatial Working Memory*. Hove: Psychology Press.
- Ma, X., and Xu, J. (2004). Determining the causal ordering between attitude toward mathematics and achievement in mathematics. *Am. J. Educ.* 110, 256–280. doi: 10.1086/383074
- MacKinnon, D. P., Krull, J. L., and Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prev. Sci.* 1, 173–181. doi: 10.1023/A:1026595011371
- MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., and Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychol. Methods* 7, 83–104. doi: 10.1037/1082-989X.7.1.83
- Maloney, E. A., Ansari, D., and Fugelsang, J. A. (2011). Rapid Communication: the effect of mathematics anxiety on the processing of numerical magnitude. *Q. J. Exp. Psychol.* 64, 10–16. doi: 10.1080/17470218.2010.533278
- Melby-Lervag, M., Redick, T. S., and Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of "far transfer": evidence from a meta-analytic review. *Perspect. Psychol. Sci.* 11, 512–534. doi: 10.1177/17456916166635612
- Menon, V. (2016). Working memory in children's math learning and its disruption in dyscalculia. *Curr. Opin. Behav. Sci.* 10, 125–132. doi: 10.1016/j.cobeha.2016.05.014
- Miller, H., and Bichsel, J. (2004). Anxiety, working memory, gender, and math performance. *Personal. Individ. Differ.* 37, 591–606. doi: 10.1016/j.paid.2003.09.029
- Nemati, P., Schmid, J., Soltanlou, M., Krimly, J.-T., Nuerk, H.-C., and Gawrilow, C. (2017). Planning and self-control, but not working memory, directly predict multiplication performance in adults. *J. Numer. Cogn.* 3, 441–467. doi: 10.5964/jnc.v3i2.61
- Ng, E., and Lee, K. (2010). Children's task performance under stress and non-stress conditions: a test of the processing efficiency theory. *Cogn. Emot.* 24, 1229–1238. doi: 10.1080/02699930903172328
- Núñez-Peña, M. I., and Suárez-Pellicioni, M. (2014). Less precise representation of numerical magnitude in high math-anxious individuals: an ERP study of the size and distance effects. *Biol. Psychol.* 103, 176–183. doi: 10.1016/j.biopsycho.2014.09.004
- Orsini, A. (1994). Corsi's block-tapping test: standardization and concurrent validity with WISC—R for Children Aged 11 to 16. *Percept. Mot. Skills* 79, 1547–1554. doi: 10.2466/pms.1994.79.3f.1547
- Owens, M., Stevenson, J., Hadwin, J. A., and Norgate, R. (2012). When does anxiety help or hinder cognitive test performance? The role of working memory capacity. *Br. J. Psychol.* 105, 92–101. doi: 10.1111/bjop.12009
- Owens, M., Stevenson, J., Norgate, R., and Hadwin, J. A. (2008). Processing efficiency theory in children: working memory as a mediator between trait anxiety and academic performance. *Anxiety Stress Coping* 21, 417–430. doi: 10.1080/10615800701847823
- Passolunghi, M. C., Caviola, S., De Agostini, R., Perin, C., and Mammarella, I. C. (2016). Mathematics anxiety, working memory, and mathematics performance in secondary-school children. *Front. Psychol.* 7:42. doi: 10.3389/fpsyg.2016.00042
- Petermann, F., Petermann, U., and Wechsler, D. (2007). *Hamburg-Wechsler-Intelligenztest für Kinder-IV: HAWIK-IV*. Göttingen: Huber.
- Ploner, N. (2014). *RON (ReadOutNumbers) [Computer Software]*. Tübingen.
- Ramirez, G., Gunderson, E. A., Levine, S. C., and Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *J. Cogn. Dev.* 14, 187–202. doi: 10.1080/15248372.2012.664593
- Roesch, S., Jung, S., Huber, S., Artemenko, C., Bahnmueller, J., Heller, J., et al. (2016). Training arithmetic and orthography on a web-based and socially-interactive learning platform. *Int. J. Educ. Inf. Technologies* 10, 204–217.
- RStudio Team (2016). *RStudio: Integrated Development Environment for R*. Boston, MA: RStudio, Inc.
- Shackman, A. J., Sarinopoulos, I., Maxwell, J. S., Pizzagalli, D. A., Lavric, A., and Davidson, R. J. (2006). Anxiety selectively disrupts visuospatial working memory. *Emotion* 6, 40–61. doi: 10.1037/1528-3542.6.1.40
- Soltanlou, M., Artemenko, C., Dresler, T., Haeussinger, F. B., Fallgatter, A. J., Ehli, A.-C., et al. (2017). Increased arithmetic complexity is associated with domain-general but not domain-specific magnitude processing in children: a simultaneous fNIRS-EEG study. *Cogn. Affect. Behav. Neurosci.* 17, 724–736. doi: 10.3758/s13415-017-0508-x
- Soltanlou, M., Artemenko, C., Ehli, A.-C., Huber, S., Fallgatter, A. J., Dresler, T., et al. (2018). Reduction but no shift in brain activation after arithmetic learning in children: a simultaneous fNIRS-EEG study. *Sci. Rep.* 8:1707. doi: 10.1038/s41598-018-20007-x
- Soltanlou, M., Pixner, S., and Nuerk, H.-C. (2015). Contribution of working memory in multiplication fact network in children may shift from verbal to visuo-spatial: a longitudinal investigation. *Front. Psychol.* 6:1062. doi: 10.3389/fpsyg.2015.01062
- Suarez-Pellicioni, M., Nunez-Pena, M. I., and Colome, A. (2016). Math anxiety: a review of its cognitive consequences, psychophysiological correlates, and brain bases. *Cogn. Affect. Behav. Neurosci.* 16, 3–22. doi: 10.3758/s13415-015-0370-7
- Thomas, G., and Dowker, A. (2000). Mathematics anxiety and related factors in young children. *Paper Presented at the British Psychological Society Developmental Section Conference*, Bristol.
- Treize, K., and Reeve, R. A. (2018). Patterns of anxiety in algebraic problem solving: a three-step latent variable analysis. *Learn. Individ. Differ.* 66, 78–91. doi: 10.1016/j.lindif.2018.02.007
- Vandierendonck, A., De Vooght, G., and Van der Gooten, K. (1998). Does random time interval generation interfere with working memory executive functions? *Eur. J. Cogn. Psychol.* 10, 413–442. doi: 10.1080/713752284
- Vandierendonck, A., Kemps, E., Fastame, M. C., and Szmalec, A. (2004). Working memory components of the Corsi blocks task. *Br. J. Psychol.* 95, 57–79. doi: 10.1348/000712604322779460
- Vukovic, R. K., Kieffer, M. J., Bailey, S. P., and Harari, R. R. (2013). Mathematics anxiety in young children: concurrent and longitudinal associations with mathematical performance. *Contemp. Educ. Psychol.* 38, 1–10. doi: 10.1016/j.cedpsych.2012.09.001
- Winer, B. J., Brown, D. R., and Michels, K. M. (1971). *Statistical Principles in Experimental Design*, Vol. 2. New York, NY: McGraw-Hill.
- Young, C. B., Wu, S. S., and Menon, V. (2012). The neurodevelopmental basis of math anxiety. *Psychol. Sci.* 23, 492–501. doi: 10.1177/0956797611429134

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

Copyright © 2019 Soltanlou, Artemenko, Dresler, Fallgatter, Ehli and Nuerk. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Effects of Working Memory, Strategy Use, and Single-Step Mental Addition on Multi-Step Mental Addition in Chinese Elementary Students

Yi Ding¹, Ru-De Liu^{2*}, Hongyun Liu², Jia Wang³, Rui Zhen⁴ and Rong-Huan Jiang²

¹ Graduate School of Education, Fordham University, New York City, NY, United States, ² Institute of Developmental Psychology, Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education, Faculty of Psychology, Beijing Normal University, Beijing, China, ³ Teachers' College, Beijing Union University, Beijing, China, ⁴ Institute of Psychological Sciences, Hangzhou Normal University, Hangzhou, China

OPEN ACCESS

Edited by:

Bert De Smedt,
KU Leuven, Belgium

Reviewed by:

Victoria Simms,
Ulster University, United Kingdom
Mojtaba Soltanlou,
University of Tübingen, Germany

*Correspondence:

Ru-De Liu
rdliu@bnu.edu.cn

Specialty section:

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

Received: 20 January 2018

Accepted: 16 January 2019

Published: 05 February 2019

Citation:

Ding Y, Liu R-D, Liu H, Wang J, Zhen R and Jiang R-H (2019) Effects of Working Memory, Strategy Use, and Single-Step Mental Addition on Multi-Step Mental Addition in Chinese Elementary Students. *Front. Psychol.* 10:148. doi: 10.3389/fpsyg.2019.00148

The aim of this paper was to examine the roles of working memory, single-step mental addition skills, and strategy use in multi-step mental addition in two independent samples of Chinese elementary students through different approaches to manipulate two dimensions of task characteristics (the primary task). In Study 1, we manipulated strategy types through the dimension of schema automaticity (whether intermediate sums were 10s) and the dimension of working memory load (WML, two steps versus four steps). A hierarchical linear model (HLM) analysis was conducted at case level, strategy level, and individual level. In Study 2, we manipulated task characteristics through schema automaticity (one-time versus two-time regrouping) and the WML (partial versus complete decomposition). A three-level HLM analysis was applied. The general findings of Study 1 and Study 2 suggested that shorter response time on single-step mental addition corresponded to shorter response time on multi-step mental addition. The use of strategies (from easier to more difficult strategies) negatively predicted response time on multi-step mental addition. Easier strategy was associated with shorter response time on multi-step mental addition. Better phonological loop was associated with shorter response time on multi-step mental addition. The findings in both studies highlighted the important role of phonological loop in mental addition in Chinese children, suggesting that the involvement of a specific subcomponent of working memory in mental arithmetic might be subject to linguistic, instructional, and contextual factors.

Keywords: working memory, automaticity, strategy use, mental addition, Chinese elementary students

INTRODUCTION

Research in mental arithmetic has received increasing attention in the past four decades (e.g., Groen and Parkman, 1972; Ashcraft, 1992, 1995; Sowder, 1992; Carroll, 1996; LeFevre et al., 2003; Liu et al., 2015). Mental arithmetic refers to the process of performing arithmetical calculation in the mind without external support such as using paper and pencil, calculators, or computers (Reys, 1984; Maclellan, 2001). Within basic arithmetic operations of addition, subtraction, multiplication, and division, addition is often learned more easily in children's learning trajectory, and addition serves as the foundation for learning the other three operations

(Beishuizen et al., 1997; Bryant et al., 1999; Torbeyns et al., 2009). In the domain of mental addition, researchers often explored simple (single-digit) mental addition and factors that affected simple mental addition in adult learners such as college students (LeFevre et al., 1996; Butterworth et al., 2001; De Rammelaere et al., 2001; Hecht, 2002). In recent years, more attention has addressed complex mental addition (e.g., addition involving two or more digits); however, the participants have been predominantly adult learners (Green et al., 2007; Imbo and LeFevre, 2009; Klein et al., 2010; Moeller et al., 2011).

Mental addition can be affected by individual characteristics such as experiences in arithmetic problem solving, working memory capacities, age, schema automaticity obtained by each individual, and strategy used for problem solving (Zbrodoff and Logan, 1986; Geary et al., 2004; Tronsky, 2005; Imbo et al., 2007; Arnaud et al., 2008). Mental addition can also be affected by task characteristics such as the difficulty level of the presented problems, types of problems (e.g., addition, subtraction, multiplication, and division), practice effects, and working memory load (WML) required by the tasks (DeStefano and LeFevre, 2004; Kalaman and LeFevre, 2007; Imbo and Vandierendonck, 2008). In our previous studies, we examined the effects of simple mental addition on complex mental addition, the effects of subcomponents of working memory on complex mental addition, and the moderating effects of working memory on single-step mental addition in relation to multi-step mental addition (Ding et al., 2017; Liu et al., 2017; Ding et al. unpublished). For our current studies, we recruited Chinese elementary students (Chinese children are anticipated to achieve a high level of proficiency of basic arithmetics in the early years of elementary school; People's Education Press, 2017) and focused on complex mental addition to explore the roles of working memory, single-step mental addition, and strategy use (manipulated by schema automaticity and WML) in multi-step mental addition.

Working Memory and Mental Arithmetic

Although children might activate different strategies for addition and multiplication, it is generally believed that children tend to be slower and make more errors with larger problems (the problem-size effect) and with problems that require carrying (DeStefano and LeFevre, 2004). If a certain amount of working memory is required for calculation of single-digit problems, we anticipate that increased working memory would be required for calculations involving multi-digit problems. Thus, in the following review of literature, we summarized findings according to single-digit problems and multi-digit problems, rather than types of calculation (i.e., addition versus multiplication).

Mental arithmetic involves encoding the presented information, executing the calculation in the mind, and providing a response (LeFevre et al., 2005). During the calculation process, one must temporarily maintain the intermediate results while continuing the calculation in order to reach the final solution. The role of working memory in mental arithmetic has been examined in empirical studies (e.g., Lemaire et al., 1996; Campbell, 1999; Lee and Kang, 2002; DeStefano and LeFevre,

2004; Meyer et al., 2010; Friso-van den Bos et al., 2013). Based on Baddeley's (1992) model of working memory, many researchers explored phonological loop, visuospatial sketchpad (VSSP), and central executive as the subcomponents of working memory in relation to mental arithmetic. However, the findings regarding the involvement of subcomponents of working memory in mental arithmetic have been quite mixed (DeStefano and LeFevre, 2004; Meyer et al., 2010; Caviola et al., 2012; Friso-van den Bos et al., 2013).

The phonological loop was found to be involved in the process of maintaining intermediate sums during multi-digit mental addition (Ashcraft and Kirk, 2001; Noël et al., 2001; DeStefano and LeFevre, 2004). Seitz and Schumann-Hengsteler (2000, 2002) found that maintaining intermediate results requires the involvement of both the central executive and the phonological loop, and they reported the involvement of the phonological loop on two-digit plus two-digit addition tasks. Heightened phonological loop skills appear to facilitate performance in complex mental addition, indicating that strong phonological loop is associated with shorter response time (Fürst and Hitch, 2000; Trbovich and LeFevre, 2003; Caviola et al., 2012).

The findings regarding the VSSP in mental arithmetic are mixed, although there was some evidence to suggest that VSSP might be involved in multi-digit problems (e.g., Logie et al., 1994; Lee and Kang, 2002; Trbovich and LeFevre, 2003; Ashkenazi et al., 2013; Laski et al., 2013). However, some studies reported null findings regarding the role of VSSP in multi-step mental arithmetic (Noël et al., 2001; Liu et al., 2017). Some reported that the impact of VSSP on mental arithmetic decreased as children matured (McKenzie et al., 2003; Holmes et al., 2008). In short, the role of the VSSP in multi-step mental arithmetic remains uncertain and warrants further research (DeStefano and LeFevre, 2004). The findings were not sufficiently comprehensive to draw a conclusion.

The central executive is responsible for planning, manipulating, and sequencing of information. The central executive also coordinates the activities of phonological loop and the VSSP. The findings regarding the central executive in mental arithmetic are inconsistent. In terms of single-digit arithmetic, some evidence pinpointed that central executive resources are required to process single-digit problems (Kaye et al., 1989; Ashcraft et al., 1992; Lemaire et al., 1996; De Rammelaere et al., 1999, 2001; Seitz and Schumann-Hengsteler, 2000, 2002; Hecht, 2002). In multi-digit arithmetic, there has been evidence for the involvement of the central executive in maintaining intermediate results during calculation (Heathcote, 1994; Logie et al., 1994; Fürst and Hitch, 2000; Seitz and Schumann-Hengsteler, 2000, 2002), whereas the influence of updating (one component of the central executive) was not significant (Liu et al., 2017).

In short, there were relevant consistent findings regarding the role of the phonological loop, rather than the VSSP and the central executive system, in mental arithmetic. In addition, because the Chinese mathematics curriculum emphasizes rote memorization, drills, and practices to enhance proficiency in mental arithmetic, the phonological loop appeared to be more relevant to the instructional and linguistic contexts in which Chinese children learn mental arithmetic. In Liu et al. (2017),

Chinese children's accuracy and response time on mental multiplication were most susceptible to phonological loop influence, when phonological loop, VSSP, and central executive tasks were tested. Thus, our examination of working memory focused on the phonological loop in the present studies.

Direct Retrieval and Schema Automaticity in Relation to Mental Arithmetic

Many factors contribute to how quickly and accurately an individual can execute mental arithmetic. One general factor is the individual's ability to understand and apply problem-solving strategies. Research has shown that it takes a long time for most students to transition from a direct modeling of the problem context, counting all of the numbers one by one, to the point that they can use direct retrieval of math facts (Mulligan and Michelmore, 1997; Downton, 2008). In comparison to low-achieving students, Zhang et al. (2014) found that high-achieving students demonstrated greater strategy flexibility during problem solving and were more accurate in direct retrieval and performing mathematics algorithm strategies. Direct retrieval is an important component of the elementary mathematics curriculum. Direct retrieval of number combinations is often achieved by typically developing students by the beginning of third grade (e.g., approximately 8 to 9 years old) in the United States (Miller and Hudson, 2007). Through repetitive practice, students learn to directly retrieve mathematics facts. Although direct retrieval is listed as one of the strategies for problem solving, it involves the retrieval of mathematics facts from long-term memory (i.e., the answer is obtained immediately) and does not involve the process of using multiple steps for problem solving.

According to the cognitive load theory (Sweller, 1988; Paas et al., 2003; van Merriënboer and Sweller, 2005), human beings have limited working memory to deal with all conscious activities and unlimited long-term memory to store facts and schemas. When students achieve automaticity with mathematics facts, they have attained a level of mastery that enables them to retrieve those facts from long-term memory without conscious effort or attention, which reflects a highly efficient process (Ponser and Snyder, 1975). A schema can be considered as a single entity that comprises multiple elements and allows humans to bypass irrelevant details. Automaticity is an important component in the process of forming schema and is often achieved after practice. In the domain of mathematics operation, an individual who has attained the level of automaticity can directly retrieve facts from long-term memory without conscious cognitive processing, which is considered direct retrieval (Siegler and Shrager, 1984; Siegler and Jenkins, 1989; Shrager and Siegler, 1998; Geary, 2011). When multiple elements of basic arithmetic facts form large operation units, students reach the level of mastery of schema automaticity after repeated practice and frequent exposure to the tasks (Sweller, 1988; Logan and Klapp, 1991; Wilkins and Rawson, 2011). For example, when a student encounters 25×6 the first time, he or she might use a regular algorithm to obtain the result. However, after repeated practice, the student might

memorize the result and directly retrieve the mathematics fact (Compton and Logan, 1991; Wilkins and Rawson, 2011) without utilizing regular operations or complex strategies.

Liu et al. (2017) reported that Chinese school systems predominantly emphasize rote memorization of single-digit and two-digit arithmetic facts. Because of repeated practice, many Chinese elementary students eventually reach the level of mastery of direct retrieval of basic arithmetic facts. Many Chinese elementary students not only retrieve basic arithmetic facts, but also rote memorize many schemas such as 25×4 and $17 + 13$. In single-step mental addition, Chinese elementary students utilize direct retrieval and schemas that become automatic. Thus, direct retrieval and schema automaticity in single-step mental addition might have an impact on the response time and accuracy rate of multi-step mental addition.

Strategy in Relation to Complex Mental Arithmetic

Children attempt different strategies such as decomposing and transformation when they solve complex arithmetic problems (Ashcraft and Fierman, 1982; Beishuizen et al., 1997; Lucangeli et al., 2003; Arnaud et al., 2008; Lemaire and Callies, 2009). For example, children could decompose " $45 + 39$ " into " $45 + 40 - 1$," " $40 + 40 + 5 - 1$," " $40 + 30 + 5 + 9$," " $45 + 30 + 9$," or " $50 + 34$." When children apply different strategies during mental arithmetic, they might utilize some schema such as " $40 + 30$ " or " $45 + 30$ " and need to activate working memory to complete processes such as transformation, temporarily memorizing intermediate sums, and operation.

When children process complex mental arithmetic, they use different strategies that are associated with different levels of schema automaticity and WML. Given an arithmetic problem (e.g., $16 + 27$), a student could use complete decomposition (e.g., decomposing $16 + 27$ to $10 + 6 + 20 + 7$, three steps in total) to carry out the calculation step by step. Step-by-step full decomposition or the use of an arithmetic algorithm often involves many steps requiring a large amount of working memory resources, which in turn might increase the response time to obtain a solution. In contrast, a student could use an automatized schema (e.g., converting $16 + 27$ to $16 + 24 + 3 = 40 + 3 = 43$, two steps in total) that leads to fewer steps (requiring fewer working memory resources) and shorter response time, in comparison to full decomposition or the use of an arithmetic algorithm. As a result, the effectiveness of a strategy used for mental arithmetic might be contingent upon the automaticity level of the strategy that was retrieved and the WML involved during problem solving.

In a previous study, we examined schema automaticity and WML through the perspective of task characteristics (Ding et al., 2017) and manipulated the levels of schema automaticity and WML (i.e., the original problem was $8 + 18 = 26$). Schema automaticity was operationalized by having the intermediate sum being 10 or the intermediate sum not being 10. In terms of WML, it was operationalized in the way that the problem had fewer versus more steps. There were four strategy conditions: (a) problems with high schema automaticity and low WML

($8 + 12 + 6 =$), (b) problems with high schema automaticity and high WML ($8 + 2 + 7 + 3 + 6 =$), (c) problems with low schema automaticity and low WML ($8 + 6 + 12 =$), and (d) problems with low schema automaticity and high WML ($8 + 6 + 3 + 7 + 2 =$). We were able to find significant main effects of schema automaticity and WML and a significant interaction effect between these two factors in mental multiplication and addition among Chinese elementary students. Our findings supported the important roles of schema automaticity and WML during mental arithmetic.

Because this study focused on Chinese children, it is important to have a brief review of how Chinese children learn math. According to Wei (2014), math education in China has a number of unique characteristics. Chinese children start learning the mathematics facts at a very young age (age 4 or 5 years through informal family education). According to People's Education Press, 2017, addition and subtraction of single-digit numbers should be mastered with high fluency by the end of first grade (age 6). Multiplication is introduced in the fall semester of second grade (age 7) and should be mastered by the end of second grade (People's Education Press, 2017). Most simple arithmetic facts such as addition, subtraction, multiplication, and division are taught through memorization and routine practice (People's Education Press, 2017). Children take at least one math class (40 min) with a single-subject math teacher (i.e., math teachers teach math classes in multiple classrooms at the same grade) each day, with at least 30 min of math homework on a daily basis. One main goal of China's math education is to develop not only conceptual understanding (what), but also procedural knowledge (how to) through practice and application (People's Education Press, 2017). Accuracy and fluency are highly regarded. From Chinese math teachers' standpoints, knowing a math concept (knowing the concept) without the abilities to efficiently solve the math problem (executing the operations) does not indicate skill acquisition. Thus, Chinese children are expected to have a very high level of accuracy and fluency on basic math facts. Given the structure of Chinese math education, automaticity and working memory appear to play a critical role in children's learning.

The Purpose of the Present Study

In Ding et al. (2017), we found significant main effects of schema automaticity and WML in relation to mental multiplication through the perspective of task characteristics (examining how the same group of students responded differently to different strategy conditions). In Liu et al. (2017), our findings indicated the important role of the phonological loop in mental multiplication through the perspective of individual characteristics (examining how individuals' subcomponents of working memory affected mental multiplication). Similar findings were revealed in our study regarding mental addition in Chinese children (2018). In short, the effectiveness of mental arithmetic is contingent upon an individual's basic mental addition skills, the strategy selected, and the working memory involved during problem solving. The purpose of this study was to examine the effects of single-step mental addition skill, strategy use, and working memory on multi-step mental addition.

Previous studies often examined the effects of simple mental arithmetic skill, strategy use, and working memory on complex mental arithmetic in isolation. We extended the previous studies in four ways. First, we simultaneously examined the effects of single-step mental addition, strategy conditions, and working memory on multi-step mental addition. Second, we manipulated the strategy through two dimensions of task characteristics, including schema automaticity and WML, to control the difficulty levels of strategy conditions. Thus, we generated four strategy conditions. We utilized the no-choice format based on Siegler and Lemaire (1997) in order to require all participants to execute the four strategies to examine how the difficulty levels of strategy use affected mental addition and this approach was validated in Ding et al. (2017). Third, we used a three-level hierarchical linear model (HLM) analysis to examine the relations of key variables at the student level, strategy level, and item level. Fourth, we tested our research questions in two studies. In Study 1 and Study 2, we used different approaches to decompose the addition problems and used different approaches to manipulate the levels of schema automaticity and WML. We wanted to explore whether Study 1 and Study 2 both supported the effects of single-step mental addition skill, strategy conditions, and working memory on multi-step mental addition. Based on the findings of Ding et al. (2017), Liu et al. (2017), Ding et al. (unpublished), we anticipated that better single-step mental addition performance would be associated with better multi-step mental addition performance (Hypothesis 1); the strategy with high schema automaticity and low WML would be associated with shorter response time on multi-step mental addition (Hypothesis 2); and better working memory capacity would be associated with shorter response time and higher accuracy rate on multi-step mental addition (Hypothesis 3) in both Study 1 and Study 2.

STUDY 1

Design

The dependent variable was the response time of the multi-step mental additions. The independent variables included the response time of the single-step mental additions, strategy conditions (we manipulated the levels of schema automaticity and WML to reflect four strategy conditions), and the phonological loop task. We considered the single-step mental addition performance as an indicator of children's basic mental addition skills. We considered the multi-step mental addition performance as an indicator of children's skills on complex mental addition.

To account for student-level, strategy-level, and item-level variances, a three-level HLM analysis was applied. At the item level (Level 1), we used multi-step mental addition performance as the dependent variable and single-step mental addition performance as the independent variable to examine the effect of single-step mental addition on multi-step mental addition. At the strategy level (Level 2), we used the four strategy conditions as the independent variable and the intercept of Level 1 as the dependent variable to examine the effects of strategy use on

multi-step mental addition. At the student level (Level 3), we used the phonological loop task as the independent variable and the intercept of Level 2 as the dependent variable to examine the effect of phonological loop on multi-step mental addition.

Measures and Procedures

Strategy

In order to manipulate the levels of schema automaticity and WML to reflect the strategy used for each question, we alternated two aspects of the structural features of addition problems: WML was manipulated by the steps involved in operations (i.e., two steps versus four steps,) and schema automaticity was manipulated by whether the single-step addition involved intermediate sums of 10 (Lemaire and Callies, 2009; Klein et al., 2010). In teaching practice, students are often taught to add base 5 numbers such as $1 + 4$, $2 + 3$, and then base 10 numbers, such as $1 + 9$, $2 + 8$, $3 + 7$, $4 + 6$, and $5 + 5$. In Chinese math curriculum, speeded arithmetic strategies are often taught to help students develop more efficient strategies and adding intermediate sums to base 10 is often utilized (e.g., transforming $7 + 9 + 13$ to $7 + 13 + 9 = 20 + 9 = 29$). Thus, in the present study, the problems with intermediate sums of 10 indicate a high level of schema automaticity, in comparison to problems without intermediate sums of 10. Given a problem such as $7 + 22 = 29$, there were four strategy conditions: (1) problems with high schema automaticity and low WML such as $7 + 13 + 9$ (there was one intermediate sum being 10 and there were two steps), (2) problems with high schema automaticity and high WML such as $7 + 3 + 4 + 6 + 9$ (there were two intermediate sums being 10 and there were four steps), (3) problems with low schema automaticity and low WML such as $7 + 9 + 13$ (there were no intermediate sums being 10 and there were two steps), and (4) problems with low schema automaticity and high WML such as $7 + 4 + 6 + 9 + 3$ (there were no intermediate sums being 10 and there were four steps) (see **Table 1**). In order to ensure the participants would perform according to the imposed problem order and format, all problems were presented in the left-to-right order.

Regression analysis treats all independent variables in the analysis as numerical, which means that these variables are interval or ratio scale variables. Our four strategy conditions

were nominal scale variables that included four categories of strategies. Thus, dummy variables were created to correctly analyze categorical variables. First, we treated the strategy condition (1) as one category and the remaining three conditions as another category. Then, we had the coding for strategy-a (3, -1, -1, -1). Second, among the strategy conditions (2), (3), and (4), we treated the condition (2) as one category, and conditions (3) and (4) as another category. Then, we obtained strategy-b (0, 2, -1, -1). Finally, we compared conditions (3) and (4), so we obtained strategy-c (0, 0, 1, -1). We did not need a fourth dummy variable to represent condition (4) because all four strategy conditions were mutually exclusive (they did not overlap) and exhaustive (no other levels existed for this variable; Ding, 2000).

Multi-Step Addition Problems (Simultaneous Presentation)

In total, there were six original questions, and each original question was presented as four strategy conditions to reflect high or low schema automaticity and high or low WML. For example, an original problem was $12 + 25$. There were four strategy conditions: (a) problems with high schema automaticity and low WML ($12 + 18 + 7 =$), (b) problems with high schema automaticity and high WML ($12 + 8 + 6 + 4 + 7 =$), (c) problems with low schema automaticity and low WML ($12 + 7 + 18 =$), and (d) problems with low schema automaticity and high WML ($12 + 4 + 7 + 6 + 8 =$). Thus, there were 24 multi-step addition problems. E-prime was used for programming. All problems were randomly presented by computers to counter the order effect. Prior to testing, a stimulus of “+” appeared in the center of the computer screen for 150 ms. The performance on addition problems measured by simultaneous presentation indicated student performance on multi-step mental addition. The participants were instructed to orally report the answer as soon as possible. When the examinee orally reported the answer, the examiner entered the answer and clicked the “return” key. Then, a stimulus of “+” appeared in the center of the computer screen and the examinee moved on to the next testing item. The computer recorded the accuracy and response time (i.e., the duration was from the point of stimulus presentation to the point that the examiner hit the enter key) for each testing item (i.e., both

TABLE 1 | Addition problems used during the testing (Study 1).

Original problems	High automaticity		Low automaticity	
	Low WML (1)	High WML (2)	Low WML (3)	High WML (4)
$8 + 18 = 26$	$8 + 12 + 6 =$	$8 + 2 + 7 + 3 + 6 =$	$8 + 6 + 12 =$	$8 + 6 + 3 + 7 + 12 =$
$12 + 25 = 37$	$12 + 18 + 7 =$	$12 + 8 + 6 + 4 + 7 =$	$12 + 7 + 18 =$	$12 + 4 + 7 + 6 + 8 =$
$21 + 25 = 46$	$21 + 19 + 6 =$	$21 + 9 + 3 + 7 + 6 =$	$21 + 6 + 19 =$	$21 + 3 + 6 + 9 + 7 =$
$24 + 27 = 51$	$24 + 26 + 1 =$	$24 + 16 + 2 + 8 + 1 =$	$24 + 1 + 26 =$	$24 + 2 + 1 + 8 + 16 =$
$17 + 17 = 34$	$17 + 13 + 4 =$	$17 + 3 + 9 + 1 + 4 =$	$17 + 4 + 13 =$	$17 + 1 + 4 + 3 + 9 =$
$19 + 35 = 54$	$19 + 31 + 4 =$	$19 + 1 + 12 + 18 + 4 =$	$19 + 4 + 31 =$	$19 + 18 + 1 + 4 + 12 =$
RT <i>M/SD</i> in seconds	3.48 (1.86)	5.28 (2.81)	6.48 (3.56)	9.03 (3.89)
Cronbach's α for RT	0.81	0.68	0.71	0.85

WML, working memory load; (1), (2), (3), and (4), conditions (1), (2), (3), and (4). RT, response time.

correct or incorrect items). Cronbach's α was 0.89 for response time and 0.69 for accuracy, which is acceptable (DeVellis, 1991).

Single-Step Addition Problems (Successive Presentation)

The same 24 addition problems were re-used. However, to obtain the accuracy and response time on single-step addition, the presentation of each testing item such as " $8 + 6 + 3 + 7 + 2 =$ " was successive. In other words, the computer first presented the single step of " $8 + 6 =$." The participant obtained the answer of "14" and pressed the "Enter" key. Then, the computer presented the next step " $+3$ "; the participant obtained the answer of "17" and pressed the "Enter" key. When the examinee orally reported the answer, the examiner entered the answer and clicked the "return" key. Then, a stimulus of "+" appeared in the center of the computer screen and the examinee moved on to the next testing item. There were 24 items in total. The response time on single-step addition was defined as the total response time on each successively presented item divided by the steps involved in that item. All problems were randomly presented by the computer. The computer automatically recorded the accuracy and response time for each item. The internal consistency for this instrument ranged from 0.68 to 0.86.

Working Memory Measure (Phonological Loop Task)

Based on our previous study examining subcomponents of working memory among Chinese elementary students, only phonological loop played a significant role in mental arithmetic, whereas VSSP and central executive did not play a significant role (Liu et al., 2017). Thus, we only included phonological loop as a measure of relevant working memory in the present study. The phonological loop task was developed based on Grant and Dagenbach (2000) and Wang et al. (2008). In total, there were 50 equations. Ten groups of equations consisted of two independent sequences of three, four, five, six, and seven equations. Participants were asked to determine whether the presented equation was correct or incorrect while they tried to memorize the second number of the equation (e.g., $7-3=4$). Both correct and incorrect answers were provided. The participants used either the "left" or the "right" button of the mouse to indicate "correct" or "incorrect." Participants were exposed to each equation a maximum of 4,000 ms. If a participant did not respond within 4,000 ms, the next equation automatically appeared on the computer. After one group of equations were presented, the participants were asked to enter all of the second numbers of those equations in a row. The E-Prime program randomly presented all equations. The second number in two adjacent equations should not be the same, and the second number in each equation should not be the same as the correct answer for that equation. The scores ranged from 0 to 50. Higher scores indicated better phonological loop. The internal consistency for this instrument in this sample was 0.81.

To counter an order effect, all problems of each task were randomly presented by the computers. E-prime was used for programming. Prior to testing, the participants received training through practice items. The participants completed three tasks in a random order.

Participants

Chinese elementary students master under-100 addition and subtraction with and without regrouping by fall semester of Grade 2. They learn under-100 multiplication and division by the end of Grade 2. Running a power analysis on a repeated measures ANOVA with four measures, a power of 0.80, an alpha level of 0.05, and a medium effect size ($f = 0.25$) requires a sample size of at least 24 (Faul et al., 2013). We recruited 40 participants for Study 1. Thus, we recruited 40 typically developing third graders who should have fluently mastered under-100 addition, subtraction, multiplication, and division by the time of testing. The average age for the participants was 8.56 years ($SD = 0.89$) and 22 were females and 18 were males. The participants were randomly recruited from an elementary school in China. This study was approved by the Research Ethics Committee of Beijing Normal University and the principals of the participating schools. Written and Informed consent was obtained from the parents/legal guardians of participants.

Results and Discussion

In Study 1, the main goal was to examine how single-step mental addition, strategy use, and working memory affected multi-step mental addition. To account for student-level, strategy-level, and item-level variances, a three-level HLM analysis was applied. Chang (2003) described HLM as a "regression of regression." The Level 3 sample size was 40, the Level 2 sample size was 160 (40 students completed four strategy conditions), and the Level 1 sample size was 960 (40 students completed all 24 items). We maintained four decimals in the HLM results because HLM results often carry very small but practically meaningful numerical values (Chang, 2003, 2004).

All of the participants had very high levels of accuracy (ranging from 84.17 to 95.71% among four conditions), and there was little variation among the participants ($M = 91.5\%$, $SD = 8.2\%$). Thus, the measure of accuracy was excluded as a variable for analysis. We only used participants' correct response

TABLE 2 | Descriptive statistics of response time at item-, strategy-, and student-level (Study 1).

Variables	N	Mean	SD	Min	Max
Item-level (Level 1)					
Multi-step RT	896	7.62	4.72	1.45	39.04
Single-step RT	896	3.69	1.85	0.86	19.59
Strategy-level (Level 2)					
Strategy-a	160				
Strategy-b	160				
Strategy-c	160				
Student-level (Level 3)					
Phono	40	36.05	8.53	10.00	47.00

We only analyzed correct response time. RT, response time (measured in seconds). There were 24 testing items. The multi-step response time was calculated based on the response time on each testing item. The single-step response time was calculated based on the response time on each testing item divided by the presentation steps involved in that item. Phono, phonological loop.

time for further analysis. The descriptive statistics of different variables are listed in **Table 2**.

In **Table 3**, the dependent variable was the average response time of multi-step addition at Level 1 (item level). The independent variable was the average response time of single-step addition at Level 1, indicating the basic single-step addition skill. γ_{100} (0.4620) was significant, which suggested that single-step response time (indicating automaticity) had an effect on multi-step response time in the positive direction. This suggested that shorter response time on single-step mental addition led to shorter response time on multi-step mental addition, supporting Hypothesis 1.

At Level 2 (strategy-level), the dependent variable was the intercept of Level 1 (the response time of multi-step mental addition). Four strategy conditions were treated as dummy variables, including strategy-a, strategy-b, and strategy-c. γ_{010} (−1.0303), γ_{020} (−0.5679), and γ_{030} (−2.2791) were all statistically significant, suggesting that strategy use had effects on multi-step response time in the negative direction. The higher the coding values for the strategies, the smaller the intercept. As we explained earlier, our dummy variables coding for strategy conditions included strategy-a (3, −1, −1, −1), strategy-b (0, 2, −1, −1), and strategy-c (0, 0, 1, −1). The values of coding of dummy variables followed a descending order from strategy (1) to strategy (4). In short, easier strategy had larger coding value and more difficult strategy had smaller coding value. The negative coefficients indicated that the strategy condition with larger coding values (an easier strategy condition) corresponded to a smaller intercept (shorter response time), whereas the strategy condition with smaller coding values (a more difficult strategy condition) corresponded to a larger intercept (longer response time). As the students moved from strategy (1) (e.g., strategy with high schema automaticity and low WML) to strategy (4)

(e.g., strategy with low schema automaticity and high WML), the intercept increased. It supported our hypothesis that the strategy with high schema automaticity and low WML would be associated with shorter response time, supporting Hypothesis 2.

At Level 3 (student-level), the independent variable was phonological loop and the dependent variable was the intercept of Level 2. The phonological loop ($\gamma_{001} = -0.1017$) negatively predicted response time on multi-step response time. As the phonological loop skill increased, the portion of intercept at Level 2 that was determined by phonological loop decreased. The higher the score on phonological loop, the lower the score on response time (shorter response time), supporting Hypothesis 3.

STUDY 2

DeStefano and LeFevre (2004) recommended that in order to further understand the role of working memory in arithmetic, researchers should systematically manipulate factors such as problem conditions, problem complexity, task requirement, and so on. Thus, it is important to manipulate task characteristics through different approaches to examine whether similar findings regarding automaticity and WML could hold true. In Study 2, the task characteristics were manipulated through the levels of schema automaticity by using one-time versus two-time regrouping and through the WML by using partial versus complete decomposition. The level of schema automaticity was manipulated through regrouping. Regrouping is defined as making groups of 10s when adding two numbers and is another name for carrying (Green et al., 2007). High schema automaticity is defined as one-time regrouping and low schema automaticity is defined as two-time regrouping. Empirical studies showed that the number of regroupings had an impact on the difficulty level

TABLE 3 | Effects of automaticity, strategy, and phonological loop on response time: three-level regression coefficients (Study 1).

Fixed effect		Coefficient	SE	T-ratio	Approx. df	P
Multi-step RT as the outcome measure						
<i>Student-level (Level 3)</i>						
For INTRCPT1	π_0					
INTRCPT2	β_{00}					
INTRCPT3	γ_{000}	9.6469	1.7847	5.41	38	<0.001
Phono	γ_{001}	−0.1017	0.0440	−2.31	38	0.027
<i>Strategy-level (Level 2)</i>						
For STRATEGY-A	β_{01}					
INTRCPT3	γ_{010}	−1.0303	0.0707	−14.58	117	<0.001
For STRATEGY-B	β_{02}					
INTRCPT3	γ_{020}	−0.5679	0.1388	−4.09	117	<0.001
For STRATEGY-C	β_{03}					
INTRCPT3	γ_{030}	−2.2791	0.2617	−8.71	117	<0.001
<i>Item-level (Level 1)</i>						
For ST-RT slope	π_1					
INTRCPT2	β_{10}					
INTRCPT3	γ_{100}	0.4620	0.1136	4.07	695	<0.001

We only analyzed correct response time. INTRCPT, intercept; Phono, phonological loop; RT, response time; ST-RT, single-step response time. It is common to retain four decimals for HLM results (Chang, 2003).

of the arithmetic problems (Imbo et al., 2007; Klein et al., 2010), which led to different levels of automatic retrieval (Siegler and Shrager, 1984; Ashcraft, 1992; Ashcraft and Christy, 1995; Hoyer et al., 2003). Problems with one-time regrouping corresponded to higher levels of schema automaticity whereas problems within two-time regrouping corresponded to lower levels of schema automaticity. The WML was manipulated through complete decomposition or partial decomposition, which led to a different number of steps in problem solving (Lemaire and Callies, 2009). In partial decomposition, only one operand was decomposed, so WML was low. In complete decomposition, two operands were both decomposed, so WML was high. Thus, we systematically manipulated the difficulty levels of automaticity and WML using different arithmetic approaches in Study 2. A similar three-level HLM analysis was utilized. If the findings in Study 1 would hold true in Study 2, it would enhance the generalization of the findings regarding the roles of automaticity and WML in mental arithmetic in Chinese students.

Design

The dependent variable was multi-step mental addition performance. The independent variables included single-step mental addition performance, strategy use (we manipulated the levels of schema automaticity and WML to reflect four strategy conditions), and phonological loop. To account for student-level, strategy-level, and item-level variances, a three-level HLM analysis was applied. At the item level (Level 1), we used single-step mental addition as the independent variable and multi-step mental addition as the dependent variable. At the strategy level (Level 2), we used the four strategy conditions as the independent variable and the intercept of Level 1 as the dependent variable. At the student level (Level 3), we used the phonological loop as the independent variable and the intercept of Level 2 as the dependent variable.

Measures and Procedures

Strategy

Similar to the design used in Study 1, we alternated two aspects of the structural features of addition problems: Schema automaticity was manipulated by the steps of regrouping involved in operations (i.e., one-time regrouping indicates high schema automaticity and two-time regrouping indicates low schema automaticity) and WML was manipulated by whether the addition involved partial decomposition (low WML) or full decomposition (high WML). There were four strategy conditions for each original question (e.g., $29 + 14 =$), consisting of (1) problems with high schema automaticity and low WML such as $(29 + 10) + 4 = ?$ (one-time regrouping and partial decomposition), (2) problems with high schema automaticity and high WML such as $(10 + 10) + (9 + 4) = ?$ (one-time and complete decomposition), (3) problems with low schema automaticity and low WML such as $(29 + 8) + 6 = ?$ (two-time regrouping and partial decomposition), and (4) problems with low schema automaticity and high WML such as $(13 + 9) + (16 + 5) = ?$ (two-time regrouping and full decomposition). See examples in **Table 4**.

Our four strategy conditions were nominal scale variables that included four categories of strategies. Thus, dummy variables were created to analyze categorical variables. First, we treated strategy condition (1) as one category and the remaining three conditions as another category. Then, we had the coding for strategy-a (3, -1, -1, -1). Second, among the strategy conditions (2), (3), and (4), we treated condition (2) as one category, and conditions (3) and (4) as another category. Then, we obtained strategy-b (0, 2, -1, -1). Finally, we compared conditions (3) and (4), so we obtained strategy-c (0, 0, 1, -1). We did not need a fourth dummy variable to represent condition (4) because all four strategy conditions were mutually exclusive (they did not overlap) and exhaustive (no other levels existed for this variable; Ding, 2000).

Multi-Step Addition Problems (Simultaneous Presentation)

First, we selected eight addition problems (the range of sums was 43 to 91, $M = 68$, $SD = 15.48$). The eight problems were designed following four rules: (a) within half of the problems, the larger operands were in the left position (e.g., $63 + 18 =$); within the other half of the problems, the larger operands were in the right position (e.g., $12 + 49 =$); (b) the digits were not repeated in the same unit or place value across operands (e.g., $64 + 14$); (c) no digits were repeated within operands (e.g., $55 + 11$); and (d) no operand had 0 in the ones place value (Lemaire and Callies, 2009).

By alternating the levels of automaticity and WML, there were four conditions for eight original problems. Thus, we had 32 problems in total. **Table 4** presents how we alternated schema automaticity and WML in the four testing conditions. Cronbach's α was 0.92 for response time and 0.67 for accuracy, which is acceptable (DeVellis, 1991). E-prime was used for programming. The details of the procedure were similar to the description in Study 1.

Single-Step Addition Problems

The same 32 addition problems were re-used. However, to obtain accuracy and response time on single-step addition, we decomposed the multi-step addition problems and generated 77 single-step addition problems. For example, $(29 + 10) + 4 =$ would be decomposed to two single-step addition problems, including $29 + 10 =$ and $39 + 4 =$. Some decomposition of the multi-step addition problems would lead to repeated single-step addition problems, and we only retained one of them. All problems were presented randomly by the computer. The stimulus of "+" was flashing in the center of the computer screen and it continued flashing for 150 ms. Then, the single-step addition problem was presented. The examinee orally reported the answer, and the examiner manually entered the answer and pushed "enter" for the next item to be presented. After the examinee completed 20 items in a row, the examinee took a short break. The computer automatically recorded the accuracy and response time (i.e., the duration was from the point of stimulus presentation to the point that the examiner hit the enter key) for each item (i.e., both correct or incorrect items). We considered the mean response time of all single-step addition

TABLE 4 | Addition problems used during simultaneous presentation and descriptive statistics (Study 2).

Original problems	High automaticity		Low automaticity	
	Low WML (1)	High WML (2)	Low WML (3)	High WML (4)
29 + 14 = 43	(29 + 10) + 4 =	(10 + 10) + (9 + 4) =	(29 + 8) + 6 =	(13 + 9) + (16 + 5) =
18 + 34 = 52	(18 + 30) + 4 =	(10 + 30) + (8 + 4) =	(18 + 26) + 8 =	(12 + 25) + (6 + 9) =
12 + 49 = 61	(10 + 49) + 2 =	(10 + 40) + (2 + 9) =	(8 + 49) + 4 =	(4 + 23) + (8 + 26) =
14 + 49 = 63	(10 + 49) + 4 =	(10 + 40) + (4 + 9) =	(6 + 49) + 8 =	(6 + 25) + (8 + 24) =
43 + 28 = 71	(40 + 28) + 3 =	(40 + 20) + (3 + 8) =	(17 + 28) + 26 =	(26 + 7) + (17 + 21) =
63 + 18 = 81	(63 + 10) + 8 =	(60 + 10) + (3 + 8) =	(24 + 18) + 39 =	(29 + 16) + (34 + 2) =
23 + 59 = 82	(20 + 59) + 3 =	(20 + 50) + (3 + 9) =	(16 + 59) + 7 =	(17 + 34) + (6 + 25) =
57 + 34 = 91	(57 + 30) + 4 =	(50 + 30) + (7 + 4) =	(57 + 16) + 18 =	(14 + 28) + (43 + 6) =
RT M/SD in seconds	5.73 (2.45)	6.02 (2.61)	11.75 (6.53)	20.30 (10.53)
Cronbach's α for RT	0.83	0.85	0.82	0.87

WML, working memory load. (1), (2), (3), and (4), conditions (1), (2), (3), and (4). Condition 1: one-time regrouping and partial decomposition. Condition 2: one-time regrouping and complete decomposition. Condition 3: two-time regrouping and partial decomposition. Condition 4: two-time regrouping and complete decomposition. RT, response time.

problems involved in a multi-step mental addition as the single-step response time corresponding to that multi-step mental addition response time.

Working Memory Measure (Phonological Loop Task)

The details of the phonological loop task were provided in Study 1.

To counter an order effect, all problems of each task were randomly presented by the computers. E-prime was used for programming. Prior to testing, the participants received training through practice items. The participants completed three tasks in a random order.

Participants

Running a power analysis on a repeated measures ANOVA with four measures, a power of 0.80, an alpha level of 0.05, and a medium effect size ($f = 0.25$) requires a sample size of at least 24 (Faul et al., 2013). We recruited 43 typically developing fourth graders (female = 25, male = 18) who should have fluently mastered under-100 addition, subtraction, multiplication, and division by the time of testing. They ranged from 9 to 11 years old ($M = 9.42$, $SD = 0.79$), with 22 females and 21 males. The participants were randomly recruited from an elementary school in China. All children did not carry documented disabilities and did not receive training on mental arithmetic. This study was approved by the Research Ethics Committee of Beijing Normal University and the principals of the participating schools. Written and Informed consent was obtained from the parents/legal guardians of participants.

Results and Discussion

In Study 2, the main goal was to examine how single-step mental addition, strategy use, and working memory measure affected multi-step mental addition. A three-level HLM analysis was applied to account for student-level, strategy-level, and item-level variances. The Level 3 sample size was 43, the Level 2 sample size was 172 (43 students completed four strategy conditions), and the Level 1 sample size was 1,375 (43 students

completed 32 items and there were missing items). Based on Chang (2003, 2004), we maintained four decimals in the HLM analysis.

All of the participants had very high levels of accuracy (89.24% for all conditions, $SD = 7.8\%$) and there was little variation among the participants. Thus, the measure of accuracy was excluded as a variable for analysis. We only used participants' correct response time for further analysis. The descriptive statistics for different variables are listed in Table 5.

In Table 6, the dependent variable was the average response time of multi-step addition at Level 1 (item level). The independent variable was the response time of single-step addition at Level 1, indicating simple addition skills. γ_{100} (1.5751) was significant and suggested that the single-step response time had an effect on multi-step response time in the positive direction. It indicated that better (faster) response time on single-step mental addition decreased the response time on multi-step mental addition, supporting Hypothesis 1.

TABLE 5 | Descriptive statistics of response time at item-, strategy-, and student-level (Study 2).

Variables	N	Mean	SD	Min	Max
Item-level (Level 1)					
Multi-step RT	1224	9.98	7.73	2.50	66.22
Single-step RT	1224	2.80	1.32	0.94	10.87
Strategy-level (Level 2)					
Strategy-a	171				
Strategy-b	171				
Strategy-c	171				
Student-level (Level 3)					
Phono	43	40.14	9.18	7.00	48.00

We only analyzed correct response time. RT, response time (measured by seconds). There were 32 testing items. The multi-step response time was calculated based on the response time on each testing item. The single-step response time was calculated based on the response time on each testing item divided by the presentation steps involved in that item. Phono, phonological loop.

TABLE 6 | Effects of automaticity, strategy, and phonological loop on response time: three-level regression coefficients (Study 2).

Fixed effect		Coefficient	SE	T-ratio	Approx. df	P
Multi-step RT as the outcome measure						
<i>Student-level (Level 3)</i>						
For INTRCPT1	π_0					
INTRCPT2	β_{00}					
INTRCPT3	γ_{000}	8.2621	1.6268	5.08	41	<0.001
Phono	γ_{001}	-0.0530	0.0257	-2.06	41	0.045
<i>Strategy-level (Level 2)</i>						
For STRATEGY-A	β_{01}					
INTRCPT3	γ_{010}	-1.3622	0.0925	-14.73	125	<0.001
For STRATEGY-B	β_{02}					
INTRCPT3	γ_{020}	-2.2642	0.1893	-11.96	125	<0.001
For STRATEGY-C	β_{03}					
INTRCPT3	γ_{030}	-4.1992	0.3914	-10.73	125	<0.001
<i>Item-level (Level 1)</i>						
For ST-RT slope	π_1					
INTRCPT2	β_{10}					
INTRCPT3	γ_{100}	1.5751	0.3050	5.16	1009	<0.001

We only analyzed correct response time. Phono, phonological loop; RT, response time; ST-RT, single-step response time; INTRCPT, intercept. It is common to retain four decimals for HLM results (Chang, 2003).

At Level 2 (strategy-level), the dependent variable was the average response time of multi-step mental addition. Four strategy conditions were treated as dummy variables, including strategy-a, strategy-b, and strategy-c. γ_{010} (-1.3622), γ_{020} (-2.2642), and γ_{030} (-4.1992) were all statistically significant, suggesting that strategy use had effects on multi-step response time in the negative direction. The negative coefficients indicated that the strategy condition with larger coding values (easier strategy condition) corresponded to a smaller intercept, whereas the strategy condition with smaller coding values (more difficult strategy condition) corresponded to a larger intercept. In other word, as students moved from strategy (1) (easier strategy) to strategy (4) (more difficult strategy), the intercept determined by the strategies increased (indicating longer response time). This supported our Hypothesis 2 that the strategy with high schema automaticity and low WML would be associated with a shorter response time.

At Level 3 (student-level), the phonological loop was the independent variable and the intercept of Level 2 was the dependent variable. The phonological loop ($\gamma_{001} = -0.0530$) negatively predicted response time on multi-step mental addition. As the phonological loop skill increased, the portion of intercept of Level 2 determined by phonological loop decreased. This finding supported our Hypothesis 3 that the higher the score of the phonological loop, the shorter the response time.

GENERAL DISCUSSION

Main Findings

The findings reveal the important roles of working memory, single-step mental addition skills, and strategy use in multi-step mental addition. We manipulated the difficulty levels of the

tasks through the dimension of WML and schema automaticity by using different approaches in Study 1 and Study 2. There are three main findings revealed in Study 1 and Study 2. First, children's shorter response time on single-step mental addition was associated with shorter response time on multi-step mental addition, regardless of how we manipulated the levels of WML and schema automaticity. Second, different strategy use was enforced through the four strategy conditions for which we manipulated the difficulty levels of schema automaticity and WML. Easier strategy was associated with shorter response time. Third, stronger phonological loop was associated with shorter response time on multi-step mental addition.

Single-step response time was considered as children's fluency on simple addition facts. Our findings support the importance of fluency in single-step addition facts in order for children to perform efficiently on multi-step mental addition. These findings confirm the importance of fluency in basic arithmetic facts, which is consistent with previous findings indicating that direct retrieval of simple mathematic facts is the most advanced and most efficient strategy with regard to problem solving speed and accuracy (Siegler, 1988; Geary et al., 2004). According to the cognitive load theory (Sweller, 1988; Paas et al., 2003; van Merriënboer and Sweller, 2005), high fluency on single-step addition largely reduces the load on working memory, freeing up working memory for more complex operations such as multi-step addition. Low fluency on single-step addition facts indicates that children who do not directly retrieve basic addition facts from their unlimited long-term memory could be overwhelmed by the number of interactive single-step addition facts that need to be processed simultaneously before multi-step addition can be processed (Paas et al., 2010). In the case that children are not fluent with single-step addition facts, their execution of single-step addition requires substantial resources of working

memory in order to consciously process the intermediate sums of single-step addition. Cumulatively, the process to execute the intermediate sums of single-step addition, memorize the intermediate sums, and add all intermediate sums to form the total sums would warrant a large amount of processing time (longer response time).

In Ding et al. (unpublished), we found that student response time followed the order of strategy (1) < strategy (2) < strategy (3) < strategy (4), from the fastest condition to the slowest condition, by examining the descriptive statistics. The findings of the HLM analysis concurred with our previous observations (Ding et al., 2017), suggesting that high schema automaticity and low WML corresponded with higher accuracy rate and shorter response time. Under the strategy (1) condition, the problems were presented with high schema automaticity and low WML such as $8 + 12 + 6$ (i.e., the difficulty levels on both dimensions were low). The problem has an intermediate sum of 10 and only has two steps. Thus, strategy (1) yielded the fastest response time. Under the strategy (4) condition, the problems were presented with low schema automaticity and high WML such as $8 + 6 + 3 + 7 + 12$ (i.e., the difficulty levels on both dimensions were high). Thus, strategy (4) yielded the slowest response time. Under strategy (2), problems with high schema automaticity and high WML ($8 + 2 + 7 + 3 + 6 =$), and strategy (3) problems with low schema automaticity and low WML ($8 + 6 + 12 =$), only one dimension of the problem was difficult and the other dimension of the problem was easy. The findings in Ding et al. (2017) indicated that sacrificing resources on WML while performing easier tasks (i.e., tasks that students could retrieve automatically) rendered shorter response time, whereas less demand on WML did not compensate for the limits of automaticity, suggesting that children performed better in condition (2) than they did in condition (3). Our Level-2 HLM analysis supported our previous observations (Ding et al., 2017). All coefficients at Level 2 are negatively significant, indicating that an easier strategy condition (condition with larger dummy variable coding value) led to a smaller intercept determined by that strategy (i.e., shorter response time). In other words, the use of a more effective strategy led to shorter response time on multi-step mental addition.

Children's performance on the phonological loop task negatively predicted the response time on multi-step mental addition, concurring with Liu et al. (2017), Ding et al. (unpublished). Higher phonological loop scores corresponded to shorter response time on multi-step mental addition. The findings underline the important role of phonological loop in mental arithmetic in Chinese children. Although there have been mixed findings regarding the role of phonological loop in single-step mental arithmetic in empirical studies conducted with Western participants (e.g., Lemaire et al., 1996; De Rammelaere et al., 1999, 2001; Seitz and Schumann-Hengsteler, 2000, 2002; Hecht, 2002), the critical role of phonological loop has been demonstrated in single-step mental arithmetic in Korean participants (Lee and Kang, 2002) and in multi-step mental arithmetic in Chinese participants (Liu et al., 2017; Ding et al. unpublished). We attributed such a universal role of phonological loop in mental arithmetic to the unique mathematics instructional approach adopted in the Chinese

education system. The Chinese school system emphasizes practice and drills on basic mathematic facts. A large amount of class time is designed to enhance children's fluency on simple arithmetic such as addition, subtraction, multiplication, and division. For example, children are required to rote memorize multiplication tables from 1×1 to 9×9 , and children often memorize such arithmetic facts through verbal rehearsal (e.g., one one equals one, one two equals two). For one-digit or two-digit addition and subtraction, rote memorization is also greatly encouraged. Thus, it is rare to observe Chinese children attempting a wide range of strategies to tackle simple arithmetic problems because they often rely on verbal modality to directly retrieve the results from long-term memory. China's Compulsory Education Law (National People's Congress, 1986) is responsible for students ranging from Kindergarten to Grade 9, and students within the age/grade range are entitled to free public education. According to the data released by the Ministry of Education (2014), there were 254,000 public schools serving students from Kindergarten to Grade 9, whereas there were only 10,425 private schools serving K-9 students (only 4% of the K-9 schools are private). In China, the standard mathematics curriculum is developed by the Ministry of Education to avoid disparities in education caused by regional differences, and all public schools (96% of all K-9 schools) utilize the standard mathematics curriculum mandated by the central government. In other words, there is very little variation in terms of how Chinese children are taught basic mathematic facts. Early mathematics teaching in China encourages language-specific representations of basic mathematic facts, which supports the critical role of phonological loop in our findings.

It is ideal to analyze findings from the aspects of accuracy and response time. However, it is noteworthy that Chinese children were fairly accurate on mental addition (91.5% accuracy rate for Study 1 and 89.24% accuracy rate for Study 2), regardless of how the testing conditions were manipulated. Thus, we did not include accuracy in the final analysis due to little variation among the students (i.e., students were fairly accurate regardless whether they spent more or less time on problems). The findings concurred with the high accuracy rate of Chinese children reported in Ding et al. (unpublished). In the present studies, we artificially increased the difficulty levels of the strategy conditions (i.e., strategies 1, 2, 3, and 4), and the complexity of the problem formats appeared to affect the response time (i.e., children took longer to respond to more complex problems). However, the Chinese children in our studies continued to accurately execute the problems and provide correct answers, regardless of the increased steps or decreased schema automaticity to retrieve arithmetic facts. The increased difficulty levels of the problems obviously sacrificed their response time, but not their accuracy rate. In China's elementary mathematics curriculum (People's Education Press, 2017), exact arithmetic calculation is largely emphasized, with less emphasis on number estimation (i.e., teachers discourage approximate answers or guessing but encourage accurate answers). A large amount of homework and in-class practice serve to enhance children's calculation accuracy. Our findings echoed the evidence of performance advances for East Asian students in simple arithmetic that occur in elementary

school as early as Kindergarten (Siegler and Mu, 2008), secondary school (Stevenson et al., 1993), and beyond (Stevenson and Stigler, 1992).

We used different approaches to alternate the difficulty levels of the strategies in Study 1 and Study 2. In Study 1, we alternated the task difficulty levels through the dimension of WML (i.e., two steps versus four steps) and the dimension of schema automaticity (i.e., intermediate sums were 10 or were not 10; Lemaire and Callies, 2009; Klein et al., 2010). Study 1 extended our previous study (Ding et al., 2017) in mental multiplication to mental addition, but followed the same design for task development. In Study 2, the schema automaticity was manipulated by the steps of regrouping involved in operations (i.e., one-time regrouping versus two-time regrouping), and WML was measured by whether the addition involved partial decomposition (low WML) or full decomposition (high WML; Lemaire and Callies, 2009), which was not utilized in previous studies. The general findings in Study 1 held true in Study 2, even though the strategy conditions were manipulated differently.

LIMITATIONS AND CONCLUSION

We note that our studies have shortcomings. First, the participants were limited to two independent samples of third graders and fourth graders in large cities of China. The findings might not be generalizable to learning of arithmetic in other countries due to possible differences in instructional approaches and learner characteristics. Second, we assumed that if a problem was presented in a specific way (e.g., imposed problem format such as $8 + 12 + 6$, then children would calculate $8 + 12 = 20$ first and then calculate $20 + 6 = 26$ in that order); that is, children would solve problems according to the enforced problem format. It remains unclear whether a small portion of participants might have generated their own strategy (e.g., $8 + 12 + 6$, then children would calculate $12 + 6 = 18$ first and then calculate $8 + 18 = 26$), regardless of the problem format we enforced. There was no mechanism to prevent spontaneous strategy use that did not follow the imposed problem format. Nevertheless, even if in some cases the students used some strategies to reorganize the sequence of calculating a problem, they must have spent some time observing the digital features of the problem and then making decisions on what strategies they could generate and use, which would have led to increased response time. Third, the measures of accuracy and response time should be used for analysis in an ideal situation. For example, both analyses for accuracy and response time were provided for the examination of mental multiplication in Ding et al. (2017). However, in the samples in the present study, students were fairly accurate on all addition task conditions regardless of how we manipulated the

tasks, although they demonstrated differentiated response time under different addition task conditions. Due to the little variance of accuracy rates among the participants, the measure of accuracy rate was excluded for final analysis.

Despite the shortcomings, the present studies extend the literature in a number of ways. First, we extended our alternation of WML and automaticity from mental multiplication (Ding et al., 2017) to mental addition. Second, within mental addition, we applied different approaches to alternate the difficulty levels of WML and schema automaticity in Study 1 and Study 2, and the general findings were consistent in both studies. Our findings indicate that future researchers might consider utilizing different approaches to alternate WML and schema automaticity and examine whether the findings hold true under different testing conditions. Third, the present studies underscore the importance of enhancing children's fluency in simple arithmetic, the use of effective strategy, and the important role of verbal representation of arithmetic facts in Chinese children. The homogenous evidence supports the activation of phonological loop during mental arithmetic problem solving in Chinese children. It highlights the importance of evaluating the linguistic features and instructional contexts in which children become fluent with basic arithmetic facts.

AUTHOR CONTRIBUTIONS

YD and RL designed the study and wrote the manuscript together. YD was in charge of the submission and review process. HL assisted for data analysis and interpretation. JW helped for data collection and data analysis. RZ and RJ assisted for editing and checking the results.

FUNDING

This study was supported by the Project of Humanities and Social Sciences Key Research Base in Ministry of Education of the People's Republic of China [教育部人文社会科学重点研究基地重大项目] (15JJD190001) to R-DL. This study was partially supported by a 2016–2017 Faculty Research Grant from Fordham University and Proof of Concept and Research Grant from the Graduate School of Education at Fordham University to YD.

ACKNOWLEDGMENTS

Thanks to the parents, teachers, and students in participated schools. Thanks to Agnes DeRaad for editorial support. Thanks to the reviewers for their helpful and constructive comments.

REFERENCES

- Arnaud, L., Lemaire, P., Allen, P., and Michel, B. F. (2008). Strategic aspects of young, healthy older adults' and Alzheimer patients' arithmetic performance. *Cortex* 44, 119–130. doi: 10.1016/j.cortex.2006.03.001
- Ashcraft, M. H. (1992). Cognitive arithmetic: a review of data and theory. *Cognition* 44, 75–106. doi: 10.1016/0010-0277(92)90051-I
- Ashcraft, M. H. (1995). Cognitive psychology and simple arithmetic: a review and summary of new directions. *Math. Cogn.* 1, 3–34.
- Ashcraft, M. H., and Christy, K. S. (1995). The frequency of arithmetic facts in elementary texts: addition and multiplication in Grades 1–6. *J. Res. Math. Educ.* 26, 396–421. doi: 10.2307/749430
- Ashcraft, M. H., Donley, R. D., Halas, M. A., and Vakali, M. (1992). "Working memory, automaticity, and problem difficulty," in *The Nature and Origins of*

- Mathematical Skills*, ed. J. I. D. Campbell (Amsterdam: Elsevier), 301–329. doi: 10.1016/S0166-4115(08)60890-0
- Ashcraft, M. H., and Fierman, B. A. (1982). Mental addition in third, fourth, and sixth graders. *J. Exp. Psychol.* 33, 216–234. doi: 10.1016/0022-0965(82)90017-0
- Ashcraft, M. H., and Kirk, E. P. (2001). The relationship among working memory, math anxiety, and performance. *J. Exp. Psychol. Gen.* 130, 224–237. doi: 10.1037/0096-3445.130.2.224
- Ashkenazi, S., Rosenberg-Lee, M., Metcalfe, A. W. S., Swigart, A. G., and Menon, V. (2013). Visuo-spatial working memory is an important source of domain-general vulnerability in the development of arithmetic cognition. *Neuropsychologia* 51, 2305–2317. doi: 10.1016/j.neuropsychologia.2013.06.031
- Baddeley, A. D. (1992). Working memory. *Science* 255, 556–559. doi: 10.1126/science.1736359
- Beishuizen, M., Van Putten, C. M., and Van Mulken, F. (1997). Mental arithmetic and strategy use with indirect number problems up to one hundred. *Learn. Instr.* 7, 87–106. doi: 10.1016/S0959-4752(96)00012-6
- Bryant, P., Christie, C., and Rendu, A. (1999). Children's understanding of the relation between addition and subtraction: inversion, identity, and decomposition. *J. Exp. Child Psychol.* 74, 194–212. doi: 10.1006/jecp.1999.2517
- Butterworth, B., Zorzi, M., Girelli, L., and Jonckheere, A. R. (2001). Storage and retrieval of addition facts: the role of number comparison. *Q. J. Exp. Psychol.* 54A, 1005–1029. doi: 10.1080/713756007
- Campbell, J. I. D. (1999). Division by multiplication. *Mem. Cogn.* 18, 245–275. doi: 10.3758/BF03198532
- Carroll, W. M. (1996). Mental computation of students in reform-based mathematics curriculum. *Sch. Sci. Math.* 96, 305–311. doi: 10.1111/j.1949-8594.1996.tb15840.x
- Caviola, S., Mammarella, I. C., Cornoldi, C., and Lucangeli, D. (2012). The involvement of working memory in children's exact and approximate mental addition. *J. Exp. Child Psychol.* 112, 141–160. doi: 10.1016/j.jecp.2012.02.005
- Chang, L. (2003). Variable effects of children's aggression, social withdrawal, and prosocial leadership as functions of teacher beliefs and behaviors. *Child Dev.* 74, 535–548. doi: 10.1111/1467-8624.7402014
- Chang, L. (2004). The role of classroom norms in contextualizing the relations of children's social behaviors to peer acceptance. *Dev. Psychol.* 40, 691–702. doi: 10.1037/0012-1649.40.5.691
- Compton, B. J., and Logan, G. D. (1991). The transition from algorithm to retrieval in memory-based theories of automaticity. *Memory Cogn.* 19, 151–158. doi: 10.3758/BF03197111
- De Rammelaere, S., Stuyven, E., and Vandierendonck, A. (1999). The contribution of working memory resources in the verification of simple mental arithmetic sums. *Psychol. Res.* 62, 72–77. doi: 10.1007/s004260050041
- De Rammelaere, S., Stuyven, E., and Vandierendonck, A. (2001). Verifying simple arithmetic sums and products: are the phonological loop and the central executive involved? *Memory Cogn.* 29, 267–273. doi: 10.3758/BF03194920
- DeStefano, D., and LeFevre, J.-A. (2004). The role of working memory in mental arithmetic. *Eur. J. Cogn. Psychol.* 16, 353–358. doi: 10.1080/09541440244000328
- DeVellis, R. F. (1991). *Scale Development: Theory and Applications*. Newbury Park, CA: Sage Publications.
- Ding, H. (2000). The approaches to analyze categorical variables in HLM [多元线性模型中定类变量的处理]. *Jiangsu Statistics [江苏统计]* 7, 21–23.
- Ding, Y., Liu, R.-D., Xu, L., Wang, J., and Zhang, D. (2017). Working memory load and automaticity in relation to mental multiplication. *J. Educ. Res.* 110, 554–564. doi: 10.1080/00220671.2016.1149794
- Downton, A. (2008). "Links between children's understanding of multiplication and solution strategies for division," in *Proceedings of the 31st Annual Conference of the Mathematics Education Research Group of Australia*, eds M. Goos, R. Brown, and K. Makar (Sydney: MERGA), 171–178.
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2013). *G* Power Version 3.1.7 [computer software]*. Kiel: Universität Kiel.
- Friso-van den Bos, I., van der Ven, S. H. G., Kroesbergen, E. H., and van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: a meta-analysis. *Educ. Res. Rev.* 10, 29–44. doi: 10.1016/j.edurev.2013.05.003
- Fürst, A. J., and Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory Cogn.* 28, 774–782. doi: 10.3758/BF03198412
- Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Dev. Psychol.* 47, 1539–1552. doi: 10.1037/a0025510
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., and DeSoto, M. C. (2004). Strategy choices in simple and complex addition: contributions of working memory and counting knowledge for children with mathematical disability. *J. Exp. Child Psychol.* 88, 121–151. doi: 10.1016/j.jecp.2004.03.002
- Grant, J. D., and Dagenbach, D. (2000). Further considerations regarding inhibitory processes, working memory, and cognitive aging. *Am. J. Psychol.* 113, 69–94. doi: 10.2307/1423461
- Green, H. J., Lemaire, P., and Dufau, S. (2007). Eye movement correlates of younger and older adults' strategies for complex addition. *Acta Psychol.* 125, 257–278. doi: 10.1016/j.actpsy.2006.08.001
- Groen, G. J., and Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychol. Rev.* 79, 329–343. doi: 10.1037/h0032950
- Heathcote, D. (1994). The role of visuo-spatial working memory in the mental addition of multi-digit addends. *Curr. Psychol. Cogn.* 13, 207–245.
- Hecht, S. A. (2002). Counting on working memory in simple arithmetic when counting is used for problem solving. *Memory Cogn.* 30, 447–455. doi: 10.3758/BF03194945
- Holmes, J., Adams, J. W., and Hamilton, C. J. (2008). The relationship between visuospatial sketchpad capacity and children's mathematical skills. *Eur. J. Cogn. Psychol.* 20, 272–289. doi: 10.1080/09541440701612702
- Hoyer, W. J., Cerella, J., and Onyper, S. V. (2003). Item learning in cognitive skill training: effects of item difficulty. *Memory Cogn.* 31, 1260–1270. doi: 10.3758/BF03195809
- Imbo, I., and LeFevre, J.-A. (2009). Cultural differences in complex addition: efficient Chinese versus adaptive Belgians and Canadians. *J. Exp. Psychol.* 35, 1465–1476. doi: 10.1037/a0017022
- Imbo, I., and Vandierendonck, A. (2008). Effects of problem size, operation, and working-memory span on simple-arithmetic strategies: differences between children and adults? *Psychol. Res.* 72, 331–346. doi: 10.1007/s00426-007-0112-8
- Imbo, I., Vandierendonck, A., and De Rammelaere, S. (2007). The role of working memory in the carry operation of mental arithmetic: number and value of the carry. *Q. J. Exp. Psychol.* 60, 708–731. doi: 10.1080/17470210600762447
- Kalman, D. A., and LeFevre, J.-A. (2007). Working memory demands of exact and approximate addition. *Eur. J. Cogn. Psychol.* 19, 187–212. doi: 10.1080/09541440600713445
- Kaye, D. B., DeWinstanley, P., Chen, Q., and Bonnefil, V. (1989). Development of efficient arithmetic computation. *J. Educ. Psychol.* 81, 467–480. doi: 10.1037/0022-0663.81.4.467
- Klein, E., Moeller, K., Dressel, K., Domahs, F., Wood, G., Willmes, K., et al. (2010). To carry or not to carry-Is this the question? Disentangling the carry effect in multi-digit addition. *Acta Psychol.* 135, 67–76. doi: 10.1016/j.actpsy.2010.06.002
- Laski, E. M., Casey, B. M., Yu, Q., Dulaney, A., Heyman, M., and Dearing, E. (2013). Spatial skills as a predictor of first grade girls' use of higher level arithmetic strategies. *Learn. Individ. Differ.* 23, 123–130. doi: 10.1016/j.lindif.2012.08.001
- Lee, K., and Kang, S. (2002). Arithmetic operation and working memory: differential suppression in dual tasks. *Cognition* 83, B63–B68. doi: 10.1016/S0010-0277(02)00010-0
- LeFevre, J. A., DeStefano, D., Coleman, B., and Shanahan, T. (2005). "Mathematical cognition and working memory," in *Handbook of Mathematical Cognition*, ed. J. I. D. Campbell (New York, NY: Psychology Press), 361–377.
- LeFevre, J. A., Sadesky, G. S., and Bisanz, J. (1996). Selection of procedures in mental addition: reassessing the problem size effect in adults. *J. Exp. Psychol.* 22, 216–230. doi: 10.1037/0278-7393.22.1.216
- LeFevre, J. A., Smith-Chant, B. L., Hiscok, K., Daley, K. E., and Morris, J. (2003). "Young adults' strategic choices in simple arithmetic: implications for the development of mathematical representations," in *The Development of Arithmetic Concepts and Skills*, eds A. J. Baroody and A. Dowker (Mahwah, NJ: Erlbaum), 203–228.
- Lemaire, P., Abdi, H., and Fayol, M. (1996). The role of working memory resources in simple cognitive arithmetic. *Eur. J. Cogn. Psychol.* 8, 73–103. doi: 10.1080/095414496383211
- Lemaire, P., and Callies, S. (2009). Children's strategies in complex arithmetic. *J. Exp. Child Psychol.* 103, 49–65. doi: 10.1016/j.jecp.2008.09.007
- Liu, R.-D., Ding, Y., Gao, B.-C., and Zhang, D. (2015). The relations between number property strategies, working memory, and multiplication in Chinese

- elementary school students. *J. Exp. Educ.* 83, 319–343. doi: 10.1080/00220973.2013.876606
- Liu, R.-D., Ding, Y., Xu, L., and Wang, J. (2017). Involvement of working memory in mental multiplication in Chinese elementary students. *J. Educ. Res.* 110, 380–390. doi: 10.1080/00220671.2015.1103689
- Logan, G. D., and Klapp, S. T. (1991). Automatizing alphabet arithmetic: I. Is extended practice necessary to produce automaticity? *J. Exp. Psychol.* 17, 179–195. doi: 10.1037/0278-7393.17.2.179
- Logie, R. H., Gilhooly, K. J., and Wynn, V. (1994). Counting on working memory in mental arithmetic. *Memory Cogn.* 22, 395–410. doi: 10.3758/BF03200866
- Lucangeli, D., Tressoldi, P. E., Bendotti, M., Bonanomi, M., and Siegel, L. S. (2003). Effective strategies for mental and written arithmetic calculation from the third to the fifth grade. *Educ. Psychol.* 23, 507–520. doi: 10.1080/0144341032000123769
- MacLellan, E. (2001). Mental calculation: its place in the development of numeracy. *Westminst. Stud. Educ.* 24, 145–154. doi: 10.1080/0140672010240205
- McKenzie, B., Bull, R., and Gray, C. (2003). The effects of phonological and visual-spatial interference on children's arithmetical performance. *Educ. Child Psychol.* 20, 93–108.
- Meyer, M. L., Salimpoor, V. N., Wu, S. S., Geary, D. C., and Menon, V. (2010). Differential contribution of specific working memory components to mathematics achievement in 2nd and 3rd graders. *Learn. Individ. Differ.* 20, 101–109. doi: 10.1016/j.lindif.2009.08.004
- Miller, S. P., and Hudson, P. J. (2007). Using evidence-based practices to build mathematics competence related to conceptual, procedural, and declarative knowledge. *Learn. Disabil. Res. Pract.* 22, 47–57. doi: 10.1111/j.1540-5826.2007.00230.x
- Ministry of Education (2014). *2014 National Education Development Statistics Bulletin*. Available at: http://www.moe.edu.cn/jyb_sjzl/s5990/201511/t20151125_220958.html
- Moeller, K., Klein, E., and Nuerk, H.-C. (2011). Three processes underlying the carry effect in addition-evidence from eye tracking. *Br. J. Psychol.* 102, 623–645. doi: 10.1111/j.2044-8295.2011.02034.x
- Mulligan, J. T., and Michelmore, M. C. (1997). Identification of multiplicative thinking in children in Grades 1-5. *J. Res. Math. Educ.* 28, 309–331. doi: 10.2307/749783
- National People's Congress (1986). *Compulsory Education Law of the People's Republic of China*. Beijing: National People's Congress.
- Noël, M. P., Désert, M., Aubrun, A., and Seron, X. (2001). Involvement of short-term memory in complex mental calculation. *Memory Cogn.* 29, 34–42. doi: 10.3758/BF03195738
- Paas, F., Renkl, A., and Sweller, J. (2003). Cognitive load theory and instructional design: recent developments. *Educ. Psychol.* 38, 1–4. doi: 10.1207/S15326985EP3801_1
- Paas, F., van Gog, T., and Sweller, J. (2010). Cognitive load theory: new conceptualizations, specifications, and integrated research perspectives. *Educ. Psychol. Rev.* 22, 115–121. doi: 10.1007/s10648-010-9133-8
- People's Education Press (2017). *Elementary Mathematics*. Beijing: People's Education Press.
- Ponser, M. I., and Snyder, C. R. (1975). "Attention and cognitive control," in *Information Processing and Cognition: The Loyola Symposium*, ed. R. L. Solso (Hillsdale, NJ: Erlbaum), 55–85.
- Reys, R. E. (1984). Mental computation and estimation: past, present, and future. *Element. Sch. J.* 84, 546–557. doi: 10.1086/461383
- Seitz, K., and Schumann-Hengsteler, K. (2000). Mental multiplication and working memory. *Eur. J. Cogn. Psychol.* 12, 552–570. doi: 10.1080/095414400750050231
- Seitz, K., and Schumann-Hengsteler, K. (2002). Phonological loop and central executive processes in mental addition and multiplication. *Psychol. Beitrage* 44, 275–302.
- Shrager, L. A., and Siegler, R. S. (1998). SCADS: a model of children's strategy choices and strategy discoveries. *Psychol. Sci.* 9, 405–410. doi: 10.1111/1467-9280.00076
- Siegler, R. S. (1988). Individual differences in strategy choices: good students, not-so-good students, and perfectionists. *Child Dev.* 59, 833–851. doi: 10.2307/1130252
- Siegler, R. S., and Jenkins, E. (1989). *How Children Discover New Strategies*. Hillsdale, NJ: Erlbaum.
- Siegler, R. S., and Lemaire, P. (1997). Older and younger adults' strategy choices in multiplication: testing predictions of ASCM using the choice/no-choice method. *J. Exp. Psychol. Gen.* 126, 71–92. doi: 10.1037/0096-3445.126.1.71
- Siegler, R. S., and Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychol. Sci.* 19, 759–763. doi: 10.1111/j.1467-9280.2008.02153.x
- Siegler, R. S., and Shrager, J. (1984). "Strategy choices in addition and subtraction: how do children know what to do?" in *Origins of Cognition Skills*, ed. C. Sophian (Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.), 229–294.
- Sowder, J. (1992). "Estimation and number sense," in *Handbook of Research on Mathematics Teaching and Learning*, ed. D. A. Grouws (New York, NY: Macmillan), 371–389.
- Stevenson, H. W., Chen, C., and Lee, S. (1993). Mathematics achievement of Chinese, Japanese, and American children: ten years later. *Science* 259, 53–58. doi: 10.1126/science.8418494
- Stevenson, H. W., and Stigler, J. W. (1992). *The Learning Gap*. New York, NY: Free Press.
- Sweller, J. (1988). Cognitive load during problem solving. *Cogn. Sci.* 12, 257–285. doi: 10.1207/s15516709cog1202_4
- Torbeyns, J., Ghesuiere, P., and Verschaffel, L. (2009). Efficiency and flexibility of indirect addition in the domain of multi-digit subtraction. *Learn. Instr.* 19, 1–12. doi: 10.1016/j.learninstruc.2007.12.002
- Trbovich, P. I., and LeFevre, J. (2003). Phonological and visual working memory in mental addition. *Memory Cogn.* 31, 738–745. doi: 10.3758/BF03196112
- Tronsky, L. N. (2005). Strategy use, the development of automaticity, and working memory involvement in complex multiplication. *Memory Cogn.* 33, 927–940. doi: 10.3758/BF03193086
- van Merriënboer, J. J. G., and Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educ. Psychol. Rev.* 17, 147–177. doi: 10.1007/s10648-005-3951-0
- Wang, E.-G., Liu, C., and Zhao, G.-X. (2008). Processing speed and working memory in children with mathematical disabilities. *Psychol. Sci.* 31, 856–860.
- Wei, K. (2014). *Explainer: What Makes Chinese Math Lessons So Good?* Available at: <http://theconversation.com/explainer-what-makes-chinese-maths-lessons-so-good-24380>
- Wilkins, N. J., and Rawson, K. A. (2011). Controlling retrieval during practice: implications for memory-based theories of automaticity. *J. Memory Lang.* 65, 208–221. doi: 10.1016/j.jml.2011.03.006
- Zbrodoff, N. J., and Logan, G. D. (1986). On the autonomy of mental processes: a case study of arithmetic. *J. Exp. Psychol. Gen.* 115, 118–130. doi: 10.1037/0096-3445.115.2.118
- Zhang, D., Wang, Q., Ding, Y., and Liu, J. (2014). Testing accommodation or modification? The effects of integrated object representation on enhancing geometry performance in children with and without geometry difficulties. *J. Learn. Disabil.* 47, 569–583. doi: 10.1177/0022219413507602

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

Copyright © 2019 Ding, Liu, Liu, Wang, Zhen and Jiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Heterosis in COMT Val158Met Polymorphism Contributes to Sex-Differences in Children's Math Anxiety

Annelise Júlio-Costa^{1,2†}, Aline Aparecida Silva Martins^{3,4†}, Guilherme Wood^{5,6}, Máira Pedrosa de Almeida^{3,4}, Marlene de Miranda^{3,4}, Vitor Gerald Haase^{1,2,5,7,8} and Maria Raquel Santos Carvalho^{3,4*}

¹ Departamento de Psicologia, FAFICH, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ² Programa de Pós-graduação em Neurociências, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ³ Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ⁴ Programa de Pós-Graduação em Genética, Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ⁵ Instituto Nacional de Ciência e Tecnologia sobre Comportamento, Cognição e Ensino (INCT-ECCE), São Carlos, Brazil, ⁶ Department of Neuropsychology, Institute of Psychology, University of Graz, Graz, Austria, ⁷ Programa de Pós-Graduação em Psicologia: Cognição e Comportamento, Departamento de Psicologia, FAFICH, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ⁸ Programa de Pós-Graduação em Saúde da Criança e Adolescente, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

OPEN ACCESS

Edited by:

Ilaria Grazzani,
University of Milano-Bicocca, Italy

Reviewed by:

Caterina Primi,
University of Florence, Italy
Ann Dowker,
University of Oxford, United Kingdom

*Correspondence:

Maria Raquel Santos Carvalho
ma.raquel.carvalho@gmail.com

[†] These authors have contributed
equally to this work

Specialty section:

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

Received: 24 February 2018

Accepted: 16 April 2019

Published: 15 May 2019

Citation:

Júlio-Costa A, Martins AAS,
Wood G, Almeida MP, Miranda M,
Haase VG and Carvalho MRS (2019)
Heterosis in COMT Val158Met
Polymorphism Contributes
to Sex-Differences in Children's Math
Anxiety. *Front. Psychol.* 10:1013.
doi: 10.3389/fpsyg.2019.01013

Math anxiety (MA) is a phobic reaction to math activities, potentially impairing math achievement. Higher frequency of MA in females is explainable by the interaction between genetic and environmental factors. The molecular-genetic basis of MA has not been investigated. The COMT Val158Met polymorphism, which affects dopamine levels in the prefrontal cortex, has been associated with anxiety manifestations. The valine allele is associated with lower, and the methionine allele with higher, dopamine availability. In the present study, the effects of sex and COMT Val158Met genotypes on MA were investigated: 389 school children aged 7–12 years were assessed for intelligence, numerical estimation, arithmetic achievement and MA and genotyped for COMT Val158Met polymorphism. The Math Anxiety Questionnaire (MAQ) was used to assess the cognitive and affective components of MA. All genotype groups of boys and girls were comparable regarding genotype frequency, age, school grade, numerical estimation, and arithmetic abilities. We compared the results of all possible genetic models: codominance (Val/Val vs. Val/Met vs. Met/Met), heterosis (Val/Met vs. Val/Val plus Met/Met), valine dominance (Val/Val plus Val/Met vs. Met/Met), and methionine dominance (Met/Met plus Val/Met vs. Val/Val). Models were compared using AIC and AIC weights. No significant differences between girls and boys and no effects of the COMT Val158Met polymorphism on numerical estimation and arithmetic achievement were observed. Sex by genotype effects were significant for intelligence and MA. Intelligence scores were higher in Met/Met girls than in girls with at least one valine allele (valine dominance model). The best fitting model for MA was heterosis. In Anxiety Toward Mathematics, heterozygous individuals presented MA levels close to the grand

average regardless of sex. Homozygous boys were significantly less and homozygous girls significantly more math anxious. Heterosis has been seldom explored, but in recent years has emerged as the best genetic model for some phenotypes associated with the *COMT* Val158Met polymorphism. This is the first study to investigate the genetic-molecular basis of MA.

Keywords: *COMT*, catechol-O-methyltransferase, heterosis, math anxiety, sex differences, dyscalculia

INTRODUCTION

Math anxiety (MA) is a learned phobic reaction toward math activities that may importantly impair math learning (Dowker et al., 2016). MA is complex and manifests itself at different levels: cognitive (negative attitudes, worrisome rumination, feelings of helplessness, low self-esteem and self-efficacy, etc.); affective (dysphoria); behavioral (avoidance, hurry to finish math tasks, etc.); and physiological (sweating, trembling, high pulse rate, etc.) (Ashcraft et al., 2007). Although MA is a multidimensional construct, it is usually measured through self-report scales focusing on two dimensions: cognitive (performance perceptions and beliefs) and affective (emotional reactions and feelings) (Wood et al., 2012, see review in Haase et al., 2019).

In this study, we investigate the relevance of the *COMT* Val158Met polymorphism for sex differences in MA. In the Introduction, we will present the following topics: (a) sex differences in MA; (b) behavioral genetics of MA; (c) genetic models; (d) *COMT* Val158Met polymorphism and cognition; (e) *COMT* Val158Met polymorphism and anxiety; (f) outline of the present study.

Sex differences in math anxiety have already been described. MA levels are significantly higher in females than in males (Hembree, 1990; Dowker et al., 2016) and in certain professional categories, such as nurses and elementary school teachers (Hembree, 1990; Beilock et al., 2010; McMullan et al., 2012). Sex differences are observed from early school age on and tend to increase over time (Dowker et al., 2012). Possible societal consequences include less participation of females in math-intensive fields (Ceci et al., 2014).

Issues involving sex, math achievement and MA are complex. Low math achievement does not seem to be the cause of higher MA levels in females. Average math performance in males and females is highly similar. In recent years, a tendency of girls to obtain better grades in math than boys has been observed (Dowker et al., 2012). However, more males than females are found at the highest levels of math performance (Wai et al., 2010; Stoet and Geary, 2013).

Some possible experiential factors associated with higher rates of MA in females would be proneness and willingness to admit anxiety symptoms (Chapman et al., 2007; McLean et al., 2011), a sex stereotype threat (Spencer et al., 1999), and social transmission of MA by female teachers (Beilock et al., 2010) and parents (Eccles et al., 1990, see review in Gunderson et al., 2012). However, higher MA levels in girls and undervaluation of girls' math abilities by parents seem to be independent of socioeconomic development and sex equity in cross-national comparisons (Stoet and Geary, 2015,

2016; Ireson, 2016). This may indicate the effects of female exposure to a more competitive environment or inherent affective/motivational differences between the sexes.

Much attention has been given to the gender stereotype threat as an important socio-cognitive mechanism underlying MA (Dowker et al., 2016). When women are reminded of the "males are better at mathematics than females" stereotype, their performance drops (Spencer et al., 1999). Neuroimaging studies indicate that the gender stereotype threat in math situations activates ventral cerebral areas associated with negative emotional processing and inhibits dorsal areas relevant to controlled and math processing (Krendl et al., 2008). However, Stoet and Geary (2012) observed that most studies only uncovered stereotype effects when prior math performance was statistically controlled. Therefore, as math performance is the outcome of interest, statistical control for prior math performance differences may confound between predictor and outcome. Stoet and Geary (2012) observed that only 55% of the studies replicated the original Spencer et al. (1999) finding, half of which adjusted for prior math achievement. Only 30% of studies without such adjustment reported significant effects of the stereotype threat.

In addition, neurocognitive differences could underlie MA sex susceptibility. This is supported by a study showing that lower MA levels in boys were mediated by better visuospatial processing abilities (Maloney et al., 2012). These subtle, but potentially relevant, cognitive differences could originate from fetal testosterone levels (Stoet and Geary, 2016). Supporting this hypothesis, a low negative correlation has been observed between 2D:4D digit-ratio, a marker of higher fetal testosterone levels, and related constructs such as math achievement and computer anxiety (de Bruin et al., 2006; Fink et al., 2006; Bull et al., 2010; Brosnan et al., 2011).

There are many hypotheses, and the origins of the higher female MA levels have been subject to considerable debate (Stoet and Geary, 2012). Overall, it is safe to conclude that both genetic and environmental factors contribute to the phenomenon. A diathesis-stress model could be advanced to explain sex differences in MA. According to this model, higher MA levels in females could be the result of interactions between specific neurocognitive vulnerabilities (such as fetal testosterone levels and yet to be discovered genetic influences) and environmental stress sources (such as low adult expectations and sex stereotype threat). Testing of this model requires a deeper understanding of the neurobiological, and especially the genetic, bases of MA. Understanding the neurogenetic underpinnings of MA susceptibility is essential for planning effective interventions.

Behavioral genetics of math anxiety have already been investigated. Two behavioral genetic studies investigated MA in twins (Wang et al., 2014; Malanchini et al., 2017). Heritability estimates were moderate (around 40%). Genetic correlations were observed with other forms of anxiety such as general anxiety and spatial anxiety. Both shared and non-shared environmental influences were uncovered. Wang et al. (2014) results suggest that MA emerges from the interaction between genetic influences on math performance and general anxiety. General anxiety, in turn, emerges from the interaction between genetic and non-shared environmental influences. Malanchini et al. (2017) obtained similar results, indicating a role for genetic and non-shared environmental factors, and for both shared and specific genetic influences on spatial anxiety and MA. No genetic or environmental sex-specific effects were investigated in these two studies.

To the best of our knowledge, no previous research has addressed the molecular-genetic underpinnings of MA. Other forms of anxiety have been associated with a host of genetic polymorphisms in several neurochemical systems (Stein et al., 2006). In this article, we focus on the dopaminergic system, as it has been implicated in various forms of performance anxiety (Mathew and Ho, 2006).

Another topic to consider is the genetic models to investigate. The impact of a specific genetic variation on a phenotype depends on the function of the protein or RNA considered. Most of the proteins are expressed from both alleles. As a consequence, the impact of genetic variants leading to amino acid substitutions that change protein function depend on the genotype, meaning the pair of alleles present on an individual. For any locus having two alleles, say, allele 1 and allele 2, the effect, the effects depend on the genotype present, 11, 12, or 22. However, it also depends on the relationship between these alleles. Consider, for example an enzyme, being 1 the wild type allele and 2 a less functional allele.

In an additive or codominance model, the genotype 11 would provide more enzyme activity, the genotype 12, less and the genotype 22 still less activity. In a 1 dominant model, 11 and 12 genotypes would produce similar enzyme function and 22 genotype would provide less (or more) enzyme activity. In the 2 dominant model, the effect would be the contrary. A third situation is seen when the both homozygous (11 and 22) genotypes produce similar enzyme activity and the heterozygous (12) genotype produces a different level of activity. When the heterozygous genotype is advantageous, the term heterosis is used. When the heterozygous genotype is disadvantageous, the term anti-heterosis is used. The term overdominance is also used, meaning heterosis. In **Figure 1**, we offer a graphic representation of these phenomena.

COMT Val158Met polymorphism has already been associated with cognition. Genetic polymorphisms in the catechol-O-methyltransferase (COMT) gene are a possible source of sex variability in cognitive and emotional processes, including math achievement and MA. The COMT Val158Met polymorphism (rs4680) has been particularly investigated. As a consequence of a nucleotide substitution in codon 158, a valine (Val) in position 158 of the protein is replaced by a methionine (Met). Three genotypes are thus defined: Val/Val, Val/Met and Met/Met,

with consequences for the enzyme's rate of catabolism. The presence of valine as compared to methionine is associated with higher COMT activity and lower dopamine availability at the synaptic cleft (Chen et al., 2004). This COMT polymorphism has been associated with several cognitive and emotional functions regulated by the prefrontal and parietal cortices, such as working memory (Goldberg et al., 2003; Mier et al., 2010; Júlio-Costa et al., 2014), numerical cognition (Tan et al., 2007; Júlio-Costa et al., 2013), impulsivity (Stein et al., 2006), anxiety (Mier et al., 2010; Gottschalk and Domschke, 2017), and psychiatric conditions such as schizophrenia (González-Castro et al., 2016), ADHD (Kebir and Joob, 2011; Bonvicini et al., 2016), autism (Nikolac Perkovic et al., 2014), etc.

Early results suggested that the valine allele would be associated with lower working memory performance and impulsivity (Stein et al., 2006; Mier et al., 2010; see also Dickinson and Elvevag, 2009). The methionine allele was, otherwise, implicated in higher working memory performance and anxiety. The connection between COMT Val158Met and numerical and arithmetic performance was explored in a study performed with typically developing children aged 7–12. The group with at least one methionine allele displayed more accurate non-symbolic number estimation (indexed by the coefficient of variation, cv), non-symbolic magnitude comparisons (indexed by the internal Weber fraction, w) and number transcoding (Júlio-Costa et al., 2013). Next, we discuss the association between the Val158Met COMT polymorphism and anxiety manifestations more specifically.

COMT Val158Met polymorphism and anxiety have also been associated. The association of the COMT Val158Met polymorphism with cognitive and emotional functions is subject to influences by culture, age and sex in adult samples (see reviews in Lee and Prescott, 2014; Barzman et al., 2015). The COMT Val158Met polymorphism has been implicated in anxiety manifestations in males and females (Hosák, 2007; Harrison and Tunbridge, 2008). Early reviews pointed out that both the valine and methionine alleles could be associated with anxiety-related phenotypes such as personality traits (e.g., neuroticism) and related disorders (e.g., generalized anxiety and panic disorder; Domschke et al., 2004; Harrison and Tunbridge, 2008). In these studies, interactions with sex were also extremely variable and complex, with a tendency for genotype-phenotype associations to be more salient in females.

More recent research also supports a nuanced picture of the association between COMT genotypes and anxiety manifestations. For example, Chen et al. (2011) found a COMT-by-sex interaction effect on affect-related personality traits in a large sample of the Chinese population. Males with at least one valine allele showed significantly higher scores on negative emotions than methionine homozygous males. Valine homozygous males presented lower scores on positive emotions, when compared with males possessing at least one methionine allele. A reverse tendency was observed in females, but the results were not significant. In another study, the Val158Met polymorphism was observed to interact with sex and neuroticism, but not with clinical symptoms of anxiety (Lehto et al., 2013). The interaction with neuroticism was investigated

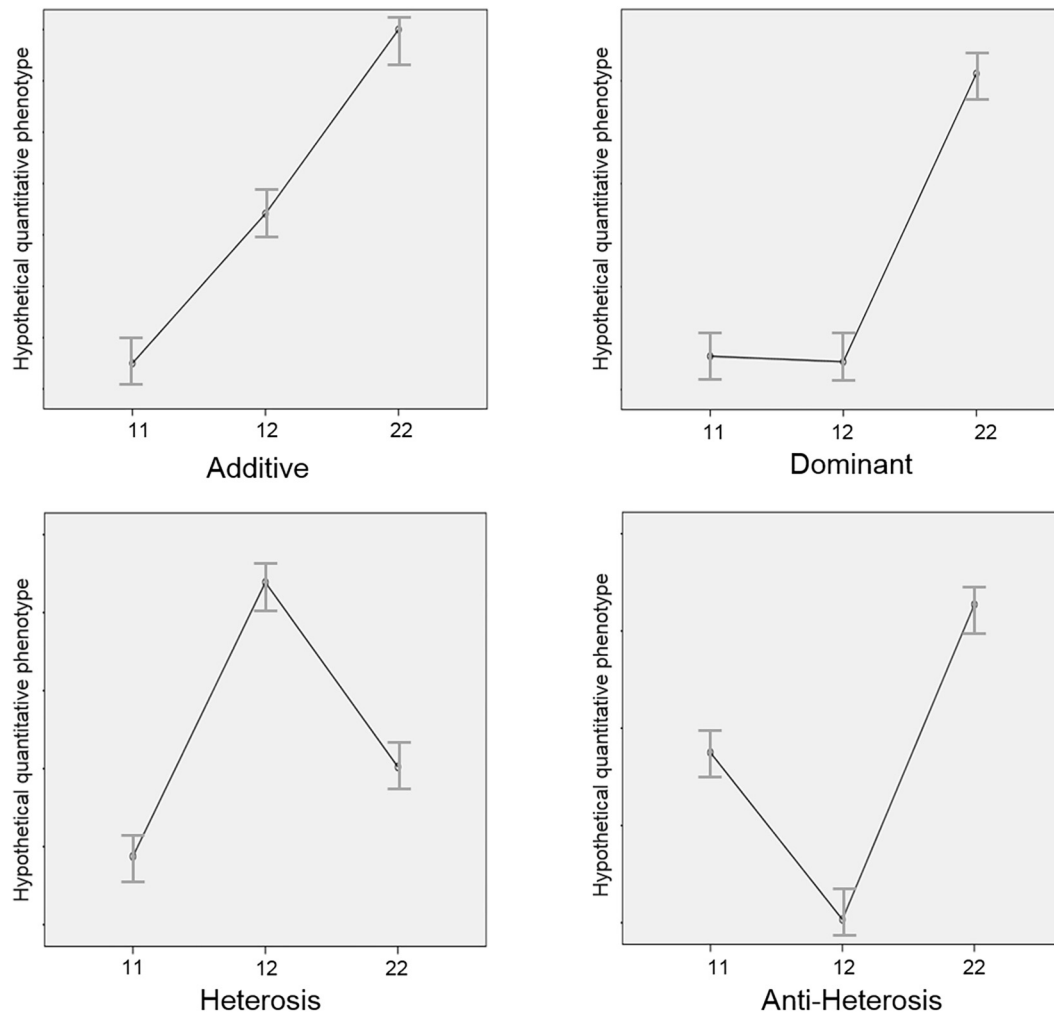


FIGURE 1 | Graphic representation of the genetic models. y-axis: any hypothetical quantitative phenotype. x-axis: 1 and 2 are different alleles at the same locus and 11, 12, and 22 are the possible genotypes. The genetic models represent the different possible interactions between alleles in a specific genotype. Starting on the top, left, and going clockwise, figures represent examples of additive (codominance), dominance, anti-heterosis and heterosis models.

at three different ages (15, 18, and 25 years) in the same cohort. Valine homozygous females presented higher levels of neuroticism in the last assessment when compared to all other sex and genotype groups. Finally, females with at least one valine allele presented a tendency for higher levels of state and trait anxiety and lower reaction times than males, when viewing faces expressing fear or anger (Domschke et al., 2012). Statistically significant higher activation rates were observed using fMRI in the ventral visual stream, amygdala, and lateral prefrontal cortex in valine homozygous females, when compared with all other sex and genotype groups. These studies show that the associations between the effects of the *COMT* Val158Met polymorphism and anxiety-related manifestations are complex and moderated by sex.

Effects of the *COMT* val158met polymorphisms may interact with sex hormones. It has been shown that estrogen down-regulates *COMT* activity; i.e., this hormone reduces the rates

of enzyme activity (Gogos et al., 1998; Xie et al., 1999; Jiang et al., 2003). Estrogen levels could then amplify the association between the valine allele and lower dopamine bioavailability in the synaptic cleft at the prefrontal cortex. A meta-analysis suggested complex interactions between the *COMT* Val158Met polymorphism and sex (Lee and Prescott, 2014). Valine homozygous males had higher neuroticism and/or harm avoidance than methionine homozygous males. No significant associations were found in women. Lee and Prescott (2014) criticize the current literature for not controlling the effects of menstrual phase and the use of hormonal birth control.

The complexity of the interactions between the *COMT* Val158Met polymorphism and sex is also reflected in studies with children and adolescents. In general, studies with children reveal that the *COMT* Val158Met polymorphism may act as a moderator between different kinds of anxiety manifestations in hetero-report measures and environmental stressors such as

early emotional trauma and maternal anxiety. Some studies have implicated the methionine allele (Olsson et al., 2007; Baumann et al., 2013) and other studies have implicated the valine allele (Sheikh et al., 2013, 2017). A dose effect for the methionine allele was observed in Olsson et al. (2007) study. The number of methionine alleles was associated with higher risk for persistent episodic anxiety in females, but not in males.

However, other studies have reported negative results, failing to find either the involvement of the Val158Met polymorphism with anxiety or the interaction with sex (Evans et al., 2009). The current state of knowledge does not allow generalizations regarding the involvement of the COMT Val158Met polymorphism in anxiety, role of the alleles involved, interactions with other genes and hormones, or interactions with sex and age. This is illustrated in **Supplementary Table S1**, which presents the methods and results of the ten original articles reporting 11 studies, identified at PubMed in October 15th, 2018 using the key words “COMT” AND “anxiety” AND “child.” Fourteen out of the 24 articles retrieved were excluded, because they were review articles, or did not investigate human subjects, did not focus on children, did not have comparison groups, or focused on psychotic and obsessive-compulsive disorder symptoms. One article reported results from two studies (Sheikh et al., 2013). Six of the 11 reported studies investigated the interaction between sex and COMT influences on anxiety.

The extant literature on effects of the sex by COMT Val158Met polymorphism on anxiety-related manifestations is scarce and extremely variable regarding age, anxiety measures, design, sampling, etc. Half of the six studies specifically examining this interaction obtained negative results. In only one of these studies with significant interactions, data were provided, from which a small effect could be estimated ($d = 0.15$) (Sheikh et al., 2017). From this literature, it is not possible to formulate more specific hypotheses on the COMT Val158Met polymorphism effects on anxiety-related manifestations that could eventually be applied to MA.

Outline of the Present Study

As reviewed above, MA is a potential cause of underrepresentation of females in math-demanding careers. According to the diathesis-stress model of etiology, MA could result from the interaction of environmental and genetic factors. Some environmental factors, such as the low expectations of parents and teachers and the stereotype threat, have been extensively investigated. Neurobiological studies have focused on the possible role of fetal testosterone levels. No previous research has directly addressed the molecular-genetic underpinnings of MA and its sex differences.

In the current study, we investigate the impact of the COMT Val158Met polymorphism and sex on numerical estimation, math achievement and MA, searching for interactions between these variables in school-age children. To this end, we genotyped the COMT Val158Met polymorphism in a group of demographically recruited, school-age children, with intelligence scores above the PR10. We also assessed the children's performance on tests of arithmetic achievement, numerical estimation and, in

MAQ, an MA self-report questionnaire (Haase et al., 2012; Wood et al., 2012).

Studies investigating the association between the COMT Val158Met polymorphism and several anxiety forms have resulted in largely incongruent and inconclusive results. A source of incongruent results in association studies is the genetic models tested. Most association studies assume codominance (additive, multiplicative, etc.) or dominance models. Heterosis has been much less frequently tested when investigating the effects of a single locus. At a single locus level, heterosis has also been referred to as molecular heterosis (Comings and MacMurray, 2000, for a review). It refers to a situation in which the phenotype in heterozygous individuals differs from that of both homozygotes. Positive heterosis refers to higher performance in heterozygotes and negative heterosis refers to lower performance in heterozygotes. Heterosis (here meaning molecular heterosis) has been frequently described for some genes expressed in the brain, including the dopamine receptors and COMT (Comings and MacMurray, 2000; Gosso et al., 2008; Luijk et al., 2011). The term overdominance is used in the literature to imply that the hybrid vigor described in association with heterosis is effectively caused by heterozygote advantage, in opposition to epigenetic effects (Charlesworth and Willis, 2009, for a review).

In the present study, we investigate four different genetic models, representing the different possible interallelic interactions in a locus. First, in the codominance model, results of the three possible genotypes (Val/Val, Val/Met, and Met/Met) are compared. Second, in the heterosis model, the results of children having the heterozygous genotype (Val/Met) are compared to a group composed of the two homozygous genotypes (Val/Val *plus* Met/Met). Third, in the valine dominance model, results from children having at least one valine allele (meaning genotypes Val/Val *plus* Val/Met) are compared with the results of children having the Met/Met genotype. Fourth, in the methionine dominance model, the results of children with at least one methionine allele (Met/Met *plus* Val/Met genotypes) are compared to the results of children having the Val/Val genotype. To the best of our knowledge, this is the first study to investigate the molecular-genetic underpinnings of MA.

MATERIALS AND METHODS

Participants

Participants were recruited from students in the 1st to 6th grades, enrolled in public and private schools in Belo Horizonte city, Brazil. Sampling was by convenience, respecting the proportion of 80% of children attending public schools, as observed in the city population. The sample covers the intermediate socio-economic strata of the Brazilian population (PR25 to PR75) (Associação Brasileira de Empresas de Pesquisa [ABEP], 2018). The sample comprised 389 children with ages ranging from 7 to 12 years (mean age = 115.66 [$sd = 12.97$] months, 55.32% female) and normal intelligence (PR > 10). Children participated only after informed consent was obtained in written form from parents and orally from themselves.

Instruments

Raven's Colored Progressive Matrices

General intelligence was assessed using the Raven's Colored Progressive Matrices – CPM (Angelini et al., 1999). *z*-scores were calculated based on the manual's norms.

Arithmetics Subtest of the Brazilian School Achievement Test (TDE)

This test is composed of three simple orally presented word problems (e.g., which is the largest, 28 or 42?) and 45 written arithmetic calculations of increasing complexity (e.g., very easy: $4-1$; easy: $1230 + 150 + 1620$; intermediate: 823×96 ; hard: $3/4 + 2/8$). Specific norms for each school grade were used to characterize children's performance (Stein, 1994; Oliveira-Ferreira et al., 2012). For the present study, the *z*-scores were calculated by grade.

Math Anxiety Questionnaire (MAQ)

The present study used a Brazilian Portuguese validated and standardized version (Haase et al., 2012; Wood et al., 2012). The MAQ items have the format of one out of four types of questions: "How good are you at..."; "How much do you like..."; "How happy or unhappy are you if you have problems with..." and "How worried are you if you have problems with...". Each question is answered in regard to six different categories related to math, namely: mathematics in general; easy calculations; difficult calculations; written calculations; mental calculations; and, math homework. Children are encouraged with supportive figures to give their responses according to a 5-point Likert scale (coded 0 to 4). Responses for each kind of question are used to build the four MAQ subscales: MAQ A – Self-perceived Performance; MAQ B – Attitudes Toward Mathematics; MAQ C – Unhappiness About Mathematics; and, MAQ D – Anxiety Toward Mathematics, according to the authors of the original British version (Thomas and Dowker, 2000). The MAQ assumes that MA is a multidimensional construct. Scales MAQ A and MAQ B assess cognitive dimensions and scales MAQ C and MAQ D tap on the affective components of MA (Wood et al., 2012). The several subscales represent correlated but independent dimensions. The best structural description reduced the MAQ to two constructs, assessing the cognitive (MAQ AB) and the affective (MAQ CD) components of MA (Wood et al., 2012). The higher the score, the higher the MA level. In the present sample, Cronbach's alpha coefficients were similar to those of the original report (Wood et al., 2012), varying from 0.76 (MAQ B) to 0.86 (MAQ Total). An age-standardized *z*-score was calculated for each MAQ scale.

Magnitude Estimation

In the non-symbolic magnitude estimation task, participants were asked to estimate, with a verbal response, the quantity of dots shown on the computer screen (Júlio-Costa et al., 2013; Pinheiro-Chagas et al., 2014). Black dots were presented in a white circle against a black background. The numerosities were 10, 16, 24, 32, 48, 56, or 64 dots. Each numerosity was presented 5 times, every time in a different configuration, such

that the same numerosity never appeared in consecutive trials. The task comprised 35 testing trials. To avoid counting, the maximum stimulus presentation time was set to 1,000 ms. As soon as the child responded, the examiner, who was seated next to the child, pressed the spacebar on the keyboard and typed the child's answer. Between individual trials, a fixation point appeared on the screen, which was a cross printed in white, with 3 cm for each line. To prevent the use of non-numerical cues, the sets of dots were generated using MATLAB in such a way that, in half of the trials, dot size remained constant and total dot area covaried positively with the numerosity; in the other half of the trials, total dot area was held constant and dot size covaried negatively with numerosity. Thus, neither total occupied area nor dot size could serve as cues for distinguishing between the different numerosities. To avoid memorization effects due to the repetition of a specific stimulus, on each trial, the stimuli were randomly chosen from a set of 10 precomputed images with the given numerosity. The data were trimmed for each subject, to exclude the responses 3 sd below or above the mean chosen value across all of the trials. As a measure of non-symbolic number representation acuity, we calculated the mean coefficient of variation (cv) of each child's responses.

Procedures

Data collection took place at the participants' schools. At first, the intelligence test (Raven's CPM) and the arithmetic subtest of the Brazilian School Achievement Test (TDE – Math) were applied in groups of eight children. This screening lasted approximately 40 min. Subsequently, parents were called to a meeting to collect the biological material (peripheral venous blood or saliva). Finally, children also answered the MAQ individually and performed the numerical magnitude estimation task in a quiet room (approximately 30 min). Data were collected from 395 children in the screening phase. Six children did not participate in the individual assessment because they performed below the PR10 on the Raven's CPM.

Genetic Analyses

DNA was extracted from peripheral blood or saliva using saline precipitation protocol (Miller et al., 1988). *COMT* Val158Met (rs4680) polymorphism was genotyped using two methods: (a) TaqMan SNP genotyping assay, genotyping was performed in ABI 7900 and analyzed using TaqMan Genotyper Software (Thermo Fisher Scientific, United States); (b) Tetra-primer amplification refractory mutation system-polymerase chain reaction (ARMS-PCR), as previously described by Ruiz-Sanz et al. (2007). In approximately 20% of the sample, genotyping was double-checked using PCR-RFLP with the restriction enzyme *Hsp92II*. This confirmed the results obtained through TaqMan SNP genotyping assay. These procedures are described in Júlio-Costa et al. (2013). Hardy-Weinberg equilibrium was tested using GenePop on the Web (Raymond and Rousset, 1995; Rousset, 2008). The predictive power sample of 80% by sex group was estimated using the Quanto software, considering an $\alpha = 0.05$ (Gauderman, 2002, 2003).

Statistical Analyses

Group differences in the distribution of sex, age, intelligence, school grade, arithmetic achievement, magnitude estimation, mathematics anxiety, as well as interactions with the *COMT* polymorphism, were examined. We explored the influence of intelligence on math achievement, numerical estimation and MA using correlation analysis, and the impact of sex using *t*-Student test. Since intelligence may confound the interpretation of possible interactions among sex, *COMT* polymorphism, school achievement and math anxiety, this variable (intelligence) was included as a covariate in further comparisons. The impact of the *COMT* polymorphism on school achievement and math anxiety was investigated by between-subjects analysis of covariance (ANCOVA).

To examine the interaction between *COMT* genotype, MA, and sex, we performed a four factorial ANCOVA using sex and *COMT* polymorphism as between-subjects factors, magnitude processing and arithmetic achievement as covariates; this procedure was repeated using each MAQ scale as the dependent variable. To test more specifically the *COMT* polymorphism effects, four different genetic models were assessed (i.e., codominance, heterosis, valine dominance, and methionine dominance). In the first model, the codominance in the *COMT* polymorphism was represented by a factor with three levels (homozygotes Val/Val, homozygotes Met/Met, and heterozygotes). In the second model, heterosis was represented by a factor with two levels (homozygotes vs. heterozygotes). In the third model, the dominance of valine was represented by a factor with two levels (Val carriers vs. non-carriers). In the fourth model, the dominance of methionine was represented by a factor with two levels (Met carriers vs. non-carriers). To establish which model accounts best for the data on MA, the Akaike Information Coefficient was calculated for each of the four models as well as the corresponding Akaike weights. A decision about the best fit was made based on the Akaike weights (Burnham and Anderson, 2003). The Akaike Information Criterion (AIC) is a simple index of the degree of disparity between a statistical model and the empirical data. The lower the value of the AIC, the better the model depicts features of the data. The AIC utilizes information on the log-likelihood of each model as well as the number of model parameters, and penalizes models with higher complexity. It is a useful tool for comparing different statistical models. When considering a set of alternative models with their respective AIC values, the Akaike weights can be calculated. These indicate, as a proportion value, how much better a model is in comparison to alternative models (see Wagenmakers and Farrell, 2004 for a primer). Statistic tests were considered significant when values of $p < 0.05$ were observed.

RESULTS

Allele frequencies observed were Met: $n = 310$ (40%) and Val: $n = 468$ (60%). Genotype frequencies for the *COMT* Val158Met polymorphism in the sample are consistent with the Hardy-Weinberg equilibrium ($p = 0.49$). Participants were assigned to one of three groups according to their genotypes: (1) homozygous

children for the valine allele (Val/Val): $n = 141$ (36.2%), (2) heterozygous children (Val/Met): $n = 186$ (47.8%) and (3) homozygous children for the methionine allele (Met/Met): $n = 62$ (15.9%). Proportions of boys and girls, their age, intelligence, grade, numerical magnitude estimation, arithmetic achievement, and MA scores are comparable [$\chi^2(1) = 0.14$; $p = 0.93$; $\eta^2 = <0.001$] across the three *COMT* genotypes (Table 1). For a predictive power of 80%, the required sample size is 147 boys and 195 girls. Our sample is composed of 174 boys and 215 girls, evidencing that the sample has enough power to detect differences in MAQ-D between the genotypic groups considering sex.

Missing data were restricted to the variable coefficient of variation (cv) of the numerical estimation task. In total, 16% of the values were missing. A chi-square test revealed that the proportion of missing values when considering sex and genotype was comparable [$\chi^2(1) = 2.26$, $p = 0.132$]. Correlation coefficients of coefficient cv (numerical estimation), arithmetic achievement and intelligence with MA were calculated (Table 2). Intelligence was positively correlated with arithmetic achievement and numerical estimation, and negatively, with MAQ B – Attitudes Toward Mathematics (one of the cognitive components of MA as assessed by MAQ). Numerical estimation correlated negatively with intelligence and arithmetic achievement, and positively with MAQ A – Self-perceived Performance. In turn, arithmetic achievement also correlated negatively with MAQ A – Self-perceived Performance, MAQ B – Attitudes Toward Mathematics, and MAQ C – Unhappiness about Mathematics. All MA subscales correlated positively with each other.

We calculated the impact of the *COMT* Val158Met polymorphism on the cv of numerical estimation, arithmetic achievement, and intelligence using ANOVA models with sex and the genetic models (i.e., codominance, heterosis, valine dominance, and methionine dominance) as between-subject factors, and compared the model fit using the AIC and AIC weights. Sex and *COMT* Val158Met polymorphism had no effect on numerical estimation, as no main- or interaction-effect reached significance (all $p > 0.2$). Sex and *COMT* Val158Met polymorphism also had no effect on arithmetic achievement (all $p > 0.3$). Importantly, an effect of *COMT* Val158Met polymorphism on intelligence was observed. The genetic model of valine dominance reached the smallest AIC ($df = 5$, AIC = 896) and the highest AIC weight (83%). All other models presented AIC values > 900 and Akaike weights $< 14\%$. In the valine dominance model, a significant interaction for sex by genotype was observed [$F(1,385) = 9.40$, $p = 0.002$, $\eta^2 = 0.023$]. None of the main-effects reached significance. Tukey *post hoc* tests revealed higher intelligence scores in Met/Met girls than in girls with at least one valine allele ($p = 0.02$).

Genetic models (i.e., codominance, heterosis, valine dominance, and methionine dominance) were compared in order to determine the contribution of valine and methionine alleles to the sex-specific phenotypes of MA. The genetic models were evaluated using four different ANCOVA models in which numerical estimation and arithmetic achievement were entered as covariates. Intelligence was not included as a covariate, for statistical reasons (Miller and Chapman, 2001), since it is also

TABLE 1 | Demographic data of children divided according to sex and genotype.

	Sex	Total n(%)	Val/Val n(%)	Val/Met n(%)	Met/Met n(%)
Sample distribution	Girls	215(55)	79 (37)	103 (48)	33 (15)
	Boys	174 (45)	62 (36)	83 (48)	29 (17)
		Mean (sd)	Mean (sd)	Mean (sd)	Mean (sd)
Grade	–	4.0 (1.0)	4.0 (0.9)	4.0 (1.0)	4.1 (1.1)
Age (months)	Girls	116 (12)	116 (13)	117 (12)	117 (13)
	Boys	115 (14)	116 (13)	114 (15)	114 (16)
Coefficient of variation in magnitude estimation (z-score)	Girls	0.19 (0.07)	0.18 (0.07)	0.19 (0.07)	0.20 (0.07)
	Boys	0.20 (0.07)	0.19 (0.06)	0.19 (0.07)	0.21 (0.06)
Intelligence (z-score)	Girls	0.71 (0.8)	0.61 (0.8)	0.60 (0.8)	0.93 (0.7)
	Boys	0.80 (0.7)	0.89 (0.7)	0.82 (0.7)	0.53 (0.9)
Arithmetic achievement (z-score)	Girls	0.29 (1.1)	0.23 (1.1)	0.19 (1.1)	0.45 (1.1)
	Boys	0.20 (1)	0.31 (1)	0.14 (1)	0.14 (1)
MAQ A – Self-perceived Performance (z-score)	Girls	0.06 (0.94)	0.11 (0.92)	0.06 (0.96)	–0.04 (0.94)
	Boys	–0.08 (1)	–0.24 (1.09)	0.04 (0.97)	–0.09 (0.79)
MAQ B – Attitudes Toward Mathematics (z-score)	Girls	0.00 (0.9)	0.05 (0.94)	–0.05 (0.91)	0.01 (0.81)
	Boys	0.01 (1.03)	–0.07 (1.01)	0.13 (1.03)	–0.13 (1.06)
MAQ C – Unhappiness About Mathematics (z-score)	Girls	.01 (0.99)	0.00 (1.11)	–0.06 (0.98)	0.27 (0.91)
	Boys	–0.02 (0.92)	0.04 (0.85)	–0.02 (0.9)	–0.11 (0.99)
MAQ D – Anxiety Toward Mathematics (z-score)	Girls	0.09 (0.99)	0.24 (1.05)	–0.04 (0.95)	0.17 (0.95)
	Boys	–0.11 (0.92)	–0.31 (0.91)	0.01 (0.83)	–0.05 (1.14)

TABLE 2 | Correlation coefficients between cognitive variables and mathematics anxiety.

N = 327	Coefficient of variation	Intelligence (z-score)	Arithmetic achievement	Attitudes Toward Mathematics	Unhappiness About Mathematics	Anxiety Toward Mathematics
Intelligence (z-score)	–0.170*					
Arithmetic achievement	–0.187**	0.414*				
MAQ A – Attitudes Toward Mathematics	0.102	–0.124*	–0.343*			
MAQ B – Unhappiness About Mathematics	0.163**	0.006	–0.198**	0.567*		
MAQ C – Anxiety Toward Mathematics	0.077	0.051	–0.018	0.317**	0.363*	
MAQ D – Self-perceived Performance	0.142**	0.068	–0.145**	0.284**	0.235**	0.556*

* $p < 0.05$, ** $p < 0.01$.**TABLE 3 |** Results and comparison of the genetic models for the COMT polymorphism on scale MAQ B – Attitudes Toward Mathematics.

Scale: MAQ B – Attitudes Toward Mathematics	Df	SS	MS	F-value	P	η_p^2	AIC/wAIC
Sex	1	0.03	0.03	0.03	0.86	0.0	910(65%)
Heterosis	1	0.54	0.54	0.60	0.45	0.0	
z-mean cv of magnitude estimation	1	9.30	9.30	10.06	0.002	0.03	
z-arithmetic achievement	1	9.02	9.02	9.75	0.002	0.03	
Sex*heterosis	1	5.20	5.20	5.62	0.02	0.02	
Residuals	321	292.35	0.93				

The other models presented AIC in the interval (913.5–913.8) and AIC weights in the interval (10.8–13%).

associated with the *COMT* Val158Met polymorphism. Model fit was compared using the AIC and AIC weights. No genetic effects were observed on MAQ A – Self-perceived Performance or MAQ C – Unhappiness about Mathematics, as no main effect of sex or genetic model effect reached significance. In contrast, the interaction of sex by genotype was significant for MAQ B – Attitudes Toward Mathematics and MAQ D – Anxiety Toward Mathematics. Attitudes Toward Mathematics was better explained by a heterosis model (Table 3). Although the interaction of sex by genotype in the heterosis model reached significance, Tukey *post hoc* comparisons did not reveal any significant difference in pairwise comparisons (all $p > 0.4$, Figure 2).

MAQ D – Anxiety Toward Mathematics was also better explained by the heterosis model (Table 4). Tukey *post hoc* comparisons revealed significant differences in pairwise comparisons between homozygous boys and girls ($p < 0.005$), but not between the heterozygous boys and girls. No other pairwise comparisons reached significance. Homozygous boys were significantly less anxious than homozygous girls, but heterozygous children were equally anxious regardless of their sex (Figure 2). MAQ D – Anxiety Toward Mathematics of heterozygous children were closer to the grand average.

DISCUSSION

In the present study, the effects of sex and *COMT* Val158Met genotypes on MA were examined in a large sample of boys and girls. No deviation from the Hardy–Weinberg Equilibrium expectancy was detected, implying that eventual differences between genotype groups do not reflect abnormalities in population genetic structure. The proportion of boys and girls in each genotype group was comparable. All genotype groups of boys and girls were comparable regarding their ages, school grades, number processing, and arithmetic abilities. Moreover, no significant differences were observed between girls and boys regarding numerical estimation or arithmetic achievement.

Intelligence was correlated positively and moderately with arithmetic achievement, and negatively and weakly with numerical estimation. Regarding MA, intelligence was negatively and weakly correlated only with the subscale MAQ B – Attitudes Toward Mathematics.

Correlations between MA and numerical/arithmetic tasks were observed. Numerical estimation correlated positively with MAQ A – Self-perceived Performance. Arithmetic achievement correlated negatively and weakly with all MA components except for MAQ D – Anxiety Toward Mathematics.

No associations between the *COMT* Val158Met polymorphism on numerical estimation and arithmetic achievement were observed. A sex by genotype interaction was observed for intelligence. Intelligence scores were higher in Met/Met girls than in girls with at least one valine allele (valine dominance model).

Our main result is related to the genetic models explaining MA. The best fitting model in both MAQ B – Attitudes Toward Mathematics and MAQ D – Anxiety Toward Mathematics was

heterosis. In the case of MAQ B – Attitudes Toward Mathematics, no *post hoc* pairwise comparisons reached significance. In contrast, in the MAQ D – Anxiety Toward Mathematics scale, homozygous boys were significantly less anxious than girls, but heterozygous children were equally anxious regardless of their sex; heterozygous individuals reported MA levels close to the grand average.

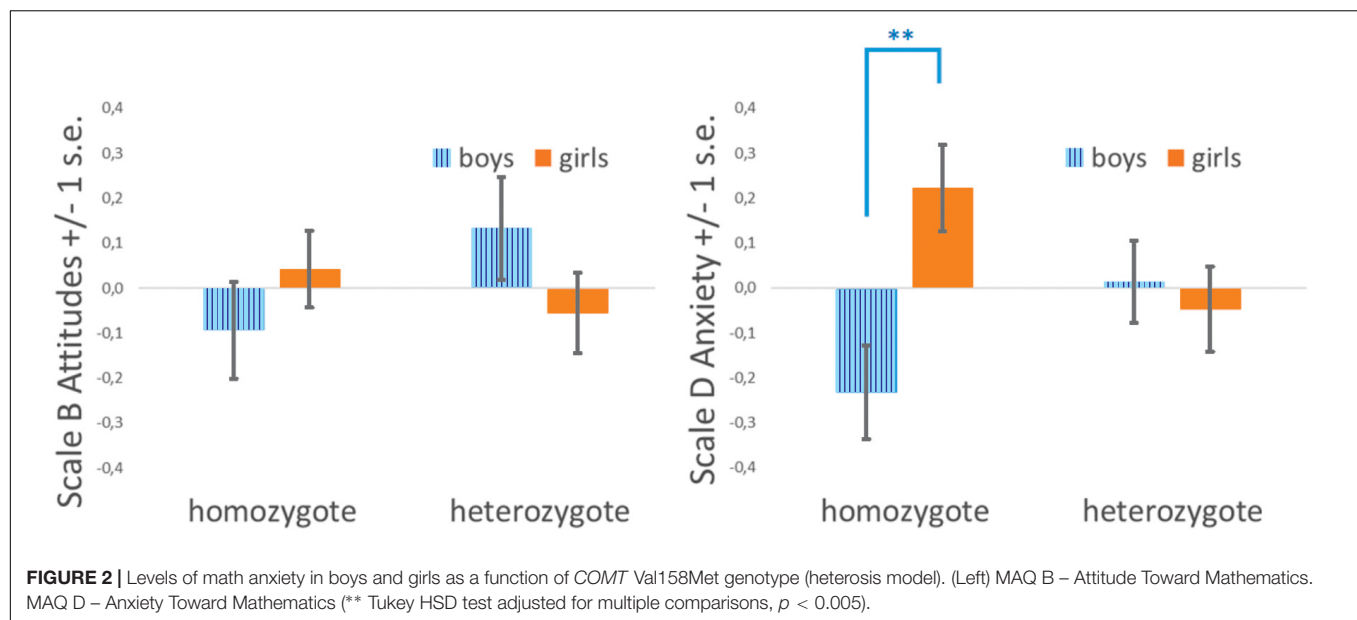
In the next sections, we discuss the validity of our results, the effects of the *COMT* Val158Met polymorphism on general and numerical cognitive measures and on MA. We conclude by discussing the importance of heterosis as an explanatory model for the effects of the *COMT* Val158Met polymorphism on several cognitive-behavioral phenotypes, including MA.

Number Processing, Arithmetic Achievement, and Intelligence

The cognitive and math-related performances of children observed in the present study were in line with data reported in the literature (Dowker et al., 2016). Intelligence was positively and moderately correlated with arithmetic achievement. Similar results have been consistently observed in other studies and intelligence is considered one of the best predictors of math achievement (Pind et al., 2003; Rohde and Thompson, 2007; Primi et al., 2010; Costa et al., 2011).

Intelligence was also negatively and weakly correlated with numerical estimation. Correlations on the same order of magnitude were observed in a large representative sample by Tosto et al. (2017). Theoretically, no correlation, or only weak correlations, might be expected, as the approximate number system (ANS) underlying numerical estimation, is usually understood to be a modular system relatively independent from general intelligence (Dehaene, 1992; Mandelbaum, 2013). However, additional evidence casts doubt on this assumption. Correlations between several tasks tapping the ANS, such as verbal estimation and symbolic and non-symbolic magnitude comparisons, are weak (Pinheiro-Chagas et al., 2014; Tosto et al., 2014). General cognitive factors (e.g., inhibitory executive functions) play a role in some ANS-related tasks, such as non-symbolic number comparison (Gilmore et al., 2013; Szűcs et al., 2013). Finally, in a large longitudinal study, age-varying patterns of predictive association were observed between prior general cognitive abilities and numerical estimation at age 16 (Tosto et al., 2017). Summing up, our results agree with the hypothesis that general cognitive requirements are important in the performance of numerical estimation tasks.

Arithmetic achievement was negatively and weakly correlated with numerical estimation, as observed by Tosto et al. (2017). Significant but small differences in verbal numerical estimation between children with and without math learning difficulties have been reported (Mejias et al., 2012; Pinheiro-Chagas et al., 2014). These results support the general view that basic numerical abilities, such as non-symbolic magnitude estimation, may be a precursor of the more advanced arithmetic abilities acquired during formal education (Piazza et al., 2010; Ferreira et al., 2012; Siegler and Braithwaite, 2017). However, the existence



and strength of these associations may vary with age, tasks and domains of math assessed (Tosto et al., 2017).

Boys and girls were comparable regarding their ages, school grades, numerical estimation and arithmetic abilities independently of their genotype groups. No significant differences were observed between girls and boys in numerical estimation or arithmetic achievement. Therefore, differences in numerical estimation or arithmetic performance cannot account for the impact of the *COMT* Val158Met polymorphism on MA.

Interestingly, higher intelligence observed in Met/Met girls yields no higher arithmetic achievement in this group which, at the first glance, seems to be counterintuitive. As discussed below, Met/Met girls have higher MA levels, which could reduce the impact of their general intellectual advantage on arithmetic achievement.

Math Anxiety

All MA subscales were positively correlated. This is in line with the literature pointing out that the four subscales of the MAQ represent different facets of the MA construct (Krinzinger et al., 2007; Wood et al., 2012), which are relatively independent from intelligence (Hembree, 1990). Accordingly, with the exception of MAQ B – Attitudes Toward Mathematics, no

MAQ subscale correlated with intelligence. MAQ B – Attitudes Toward Mathematics exhibited a weak negative correlation with intelligence, which corroborates previous findings (Minato and Yanase, 1984; Moenikia and Zahed-Babelan, 2010) since, in the MAQ B scale, higher scores code for more negative attitudes toward mathematics.

Arithmetic achievement was negatively and weakly correlated with all MAQ scales, except for MAQ D – Anxiety Toward Mathematics. These results are also in line with previous studies (Moenikia and Zahed-Babelan, 2010). Since correlations between arithmetic achievement and MA are more pronounced in the subscale measuring the affective component of MA (Krinzinger et al., 2007; Haase et al., 2012; Wood et al., 2012), the effects of sex by *COMT* genotype interactions on MA seem to be emotionally mediated.

Our study was not designed to answer the question of the specificity of results regarding MA, as we did not use measures of more generalized anxiety or reading/spelling performance. MA is a complex construct, including both cognitive and affective dimensions (Dowker et al., 2016; Haase et al., 2019). Behavioral genetic models have shown that MA shares considerable sources of genetic and environmental influences with other anxiety-related constructs (Wang et al.,

TABLE 4 | Results and comparison of the genetic models for the *COMT* polymorphism on scale MAQ D – Anxiety Toward Mathematics.

Scale: Maq D – Anxiety toward Mathematics	Df	SS	MS	F-value	p	η_p^2	AIC/wAIC
Sex	1	3.02	3.02	3.30	0.07	0.0	907 (72%)
Met dominance	1	0.01	0.01	0.01	0.91	0.003	
z-mean cv of magnitude estimation	1	6.37	6.37	6.96	0.01	0.02	
z-arithmetic achievement	1	5.00	5.00	5.46	0.02	0.02	
Sex*Met dominance	1	11.20	11.20	12.24	0.0005	0.04	
Residuals	321	293.74	0.915				

The other models presented AIC in the interval (910–918) and AIC weights in the interval (0–15%).

2014; Malanchini et al., 2017). However, correlations between MA and other forms of anxiety are usually weak ($r = 0.3$) (Hembree, 1990), suggesting that MA and other forms of anxiety represent partially independent dimensions. In a previous study using MAQ in school-aged children, we observed that correlations with generalized anxiety (assessed by CBCL) were weak, and that MAQ levels were associated with math performance but not with word spelling performance (Haase et al., 2012). The reverse pattern was observed for generalized anxiety. Generalized anxiety was associated with spelling but not with math performance. Considering the behavioral genetic results, it is safe to conclude that the construct MA refers to the content of phobic reactions in predisposed individuals.

Finally, differences in the covariance structure of MA in children with different genotypes are possible but remain elusive in the present study. This is because our sample size is not large enough for a useful estimation of correlations coefficients for different groups separately, particularly when considering only the boys or only the girls with the Met/Met genotype.

COMT Val158Met Polymorphism and Cognition

No main or interaction effects of the factors sex and *COMT* polymorphism on basic magnitude estimation or arithmetic achievement were observed in the present study.

A link between dopaminergic activity and magnitude processing was established in experimental research in rodents. In rodents, pharmacological inhibition or facilitation of dopaminergic activity modulates temporal and numerical magnitude estimation (Cordes et al., 2007; Coull et al., 2011). Dopaminergic activity is related to the speed of the counting mechanism underlying magnitude estimation according to the accumulator model (Leslie et al., 2007). In humans, one study from our research group investigated the impact of the *COMT* Val158Met polymorphism on basic number processing tasks (Júlio-Costa et al., 2013). In that study, children with at least one methionine allele presented better performance in the numerical estimation and other numerical and arithmetic tests. The discrepancy between that study and the present one is only apparent. A large proportion of the sample assessed by Júlio-Costa et al. (2013) was also included in the present study. Therefore, disappearance of the effect with increase of sample size is indicative of a false positive result, probably caused by the smaller sample investigated in that study. The sample size of 327 children, for whom data were available on cv in the current report, offers a higher degree of protection against false positive findings and may be given more weight than the partial evidence published previously. Accordingly, evidence for a detectable impact of the *COMT* Val158Met polymorphism on basic magnitude estimation remains elusive, since the positive evidence obtained in rodents using pharmacological manipulations are much stronger than the functional differences occurring naturally between the valine and methionine containing enzyme.

Beyond the scope of basic magnitude estimation, sex and *COMT* Val158Met polymorphism also seem to have no impact on arithmetic achievement. In a small study using fMRI, Tan et al. (2007) explored the role of the *COMT* Val158Met genotypes in numerical/arithmetic processing. Adult carriers of the valine allele had higher levels of dorsolateral prefrontal cortex activation than individuals with other genotypes. This activation correlated with arithmetic operations that require working memory, but not with the operations requiring long-term memory retrieval. The increased brain activation during resolution of arithmetic problems in individuals with the valine allele may be interpreted as a compensatory mechanism (Tan et al., 2007). Consistent with the present study, however, no effects of genotype were observed at the behavioral level.

The connection between the *COMT* Val158Met polymorphism and numerical/arithmetic performance could also be investigated in 22q11.2 microdeletion syndrome (22q11.2DS). Individuals with 22q11.2DS present several phenotypic traits such as risk of schizophrenia, intellectual disability and math learning difficulties in the presence of hemizygosity at the *COMT* Val158Met locus (Karayiorgou et al., 2010). Some research supports a role for the valine allele in intellectual disability and schizophrenia (Shashi et al., 2006, 2010), but results have not always been replicated (Campbell et al., 2010; Franconi et al., 2016). However, to the best of our knowledge, the specific association between *COMT* Val158Met polymorphism and numerical/arithmetic abilities has not yet been investigated in 22q11.2DS.

A sex by genotype interaction was detected for intelligence. Met/Met girls exhibited higher intelligence scores compared to girls with at least one valine allele (valine dominance). The methionine allele is associated with higher intelligence in some studies (Enoch et al., 2009; Carmel et al., 2014), higher cognitive performance, and also higher anxiety levels (Stein et al., 2006; Dickinson and Elvevag, 2009). Specifically, Ramirez et al. (2013) showed that MA is higher on the extremes of the distribution of working memory capacity. Since Met/Met girls generally present higher cognitive ability, they would also be more affected by MA, as observed in the present study. One possible mechanism of how higher levels of MA may impair arithmetic achievement has been proposed by Ramirez et al. (2013). According to them, high performing individuals tend to rely on working memory-intensive solution strategies, which are likely disrupted when MA interferes with working memory. Therefore, Met/Met girls have high levels of intelligence but also high MA levels, which could reduce their general intellectual advantage on arithmetic achievement. These findings suggest a sex-specific connection between higher cognitive abilities, the Met/Met genotype, and susceptibility to interference of MA on math performance, which will be explored below in further detail. However, this connection between high intelligence and MA in girls is not the whole story, since Val/Val girls were not more intelligent than other groups of children in our study. Interestingly, evidence indicates higher levels of neuroticism in Val/Val women from adolescence to

young adulthood (Lehto et al., 2013). Therefore, there is evidence of higher levels of anxiety in both homozygous genotypes of female participants. These pieces of evidence also will be discussed in further detail in relation to the heterosis model in the next section.

COMT Val158Met Polymorphism and MA

Sex and *COMT* polymorphism had a marginal effect on subscale MAQ B – Attitudes Toward Mathematics, which represents a more cognitive aspect of MA. Here, it is important to consider the young age of the participants in our study. It is possible that their self-concept and attitudes toward mathematics were not yet as fully developed as later in puberty and adulthood. Stronger effects of sex on the cognitive aspect of MA are known to become more evident in older adolescents (Utsumi and Mendes, 2000; Mata et al., 2012). A specific interaction between sex and grade was obtained by Wigfield and Meece (1988). These authors assessed the cognitive and affective dimensions of MA in 564 children from 6th to 12th grades. Grade differences were observed only in the cognitive dimension, with older children scoring higher than younger ones. Sex differences were observed only in the affective dimension of MA, with girls scoring higher. No sex by grade interactions were observed (Wigfield and Meece, 1988).

In our study, more robust effects were observed in the subscale MAQ D – Anxiety Toward Mathematics, which represents a more affective aspect of MA. The interaction between sex and the *COMT* Val158Met genotype in MAQ D – Anxiety Toward Mathematics was significant under the heterosis model. Significant differences between boys and girls were observed in both homozygous groups, but not in heterozygous individuals. Homozygous girls presented higher levels of MA than boys, while heterozygous boys and girls did not differ regarding MAQ D – Anxiety Toward Mathematics. The existing literature on sex-related differences in anxiety levels associated with the Met/Met and Val/Val genotypes (reviewed in **Supplementary Table S1**) suggests different explanations for the higher levels of MA observed in Met/Met or in Val/Val girls. Our comparison of genetic models suggests that these apparently contradictory results may reflect the fact that the heterosis model has not been tested. The bulk of the literature on the *COMT* Val158Met polymorphism focuses on statistical models separating all three genotypes (codominance or additive models) or genotypes organized in two groups (dominance models). Therefore, cases in which heterosis is the correct genetic model for the data may have been easily overlooked; and, the number of explanations for the phenotypes connected with the different genotypes may be artificially inflated.

The genetic heterosis model of MAQ D – Anxiety Toward Mathematics suggests that homozygous girls are more susceptible than boys to the emotional arousal elicited by math tasks perceived as difficult. Whether the causal pathways of the genetic effects of Met/Met and Val/Val genotypes are the same or not is an open question. This can be answered only with more detailed studies. In the final two sections, we are going to discuss: (a) the mechanisms of estrogen effects that may contribute to increase the levels of MAQ D – Anxiety Toward Mathematics

in homozygous girls; and (b) the role of heterosis in the *COMT* Val158Met polymorphism.

Mechanisms of Estrogen Effects on COMT

Sex differences in many behavioral traits in humans have been described and attributed to the influence of sex hormones through their influences on neurotransmitter systems, such as dopamine (Sherwin, 2007; Riccardi et al., 2011; for an overview of dopamine system, see Wahlstrom et al., 2010).

The increase in the estrogen levels during puberty down-regulates *COMT* transcription and leads to sex differences in *COMT* enzyme activity (Xie et al., 1999; Tunbridge, 2010). As a consequence, females show higher dopamine levels in the synaptic cleft in brain regions where *COMT* is the main metabolizer of dopamine. Here, it is important to remember that dopamine can be depleted from the synaptic cleft by *DAT1* or by *COMT*. Consequences of *COMT* malfunctioning are more prominent in those regions where *DAT1* is physiologically poorly expressed, such as the prefrontal cortex. Therefore, functional *COMT* polymorphisms have a strong impact in cognitive tasks associated with attention and executive functions (Riccardi et al., 2006, 2011).

There is another interaction mechanism between *COMT* and estrogen. *COMT* metabolizes catechol estrogens (i.e., 2-OHE2, 2-OHE1, 4-OHE2, and 4-OHE1) to methyl-estrogen, which has been associated with cancer development and progression (Dawling et al., 2001; Ashton et al., 2006). These pathways have not yet been investigated in relation to cognitive functions.

The relationship between estrogen and *COMT* is even more complex. Men have 17% higher *COMT* activity in the prefrontal cortex than women, independently of any polymorphisms (Chen et al., 2004). Higher *COMT* activity in men has been described in most tissues (reviewed by Harrison and Tunbridge, 2008). Evidence for sexual dimorphism in *COMT* associated phenotypes is abundant, but frequently conflicting, in the literature. This suggests that a “dopaminergic tonus” or “optimal dopamine level” may differ according to sex, age, brain region or system, physiological or pathological state as well as pharmacological responses (Jacobs and D’Esposito, 2011).

These effects are clear in the evaluation of the impact of the *COMT* Val158Met polymorphism. For example, the association of the Met allele with obsessive-compulsive disorder in men but not in women (Karayiorgou et al., 1997) is a well replicated finding (Pooley et al., 2007). The Met allele has been associated with increased levels of anxiety and cautious personality (Enoch et al., 2003; Olsson et al., 2005; Montag et al., 2012) (see **Supplementary Table S1**), and anxiety disorder (Domschke et al., 2004; Woo et al., 2004; Rothe et al., 2006).

However, it is important to consider that physiological sex differences occur in many systems, not only in sex hormones. As a consequence, effects attributed to differences in sex hormones may reflect differences in other, less investigated biological systems.

COMT genotype and sex hormone influence may interact epigenetically in complex ways. In this sense, it is also important

to consider that our participants were prepubertal children. Investigations with post-pubertal participants should follow the recommendations of Lee and Prescott (2014) to consider menstrual cycle phase and use of oral contraceptives.

Evidence for Heterosis in the *COMT* Val158Met Polymorphism

As mentioned above, the interaction between the alleles may take four main forms: codominance, heterosis, Val dominance and Met dominance. The term codominance is used to refer to situations in which the three genotypes have different effects at the phenotype level. The most used codominance model is the additive, in which each allele substitution has a incremental effect (e.g., considering a locus with two alleles, 1 and 2, the effects of the genotypes would be $11 < 12 < 22$). In the heterosis model, heterozygous individuals have a phenotype that differs from both homozygous groups, which have similar phenotypes. The phenotype in heterozygous individuals can be advantageous (positive heterosis) or disadvantageous (negative heterosis).

Several examples of heterosis in the *COMT* Val158Met polymorphism have been reported already. We will discuss only those studies with substantial sample sizes. Barnett et al. (2007) investigated the effects of the *COMT* Val158Met polymorphism on working memory, verbal and motor inhibition, attentional control, and IQ in a sample composed of 8,707 children, aged 8–10 years. These authors described heterozygous advantage in a measure of sustained attention in boys but not in girls.

Gosso et al. (2008) described an example of positive heterosis in working memory. These authors investigated a sample of over 600 participants, approximately half of them children. Positive heterosis was detected: better results in working memory tests were found in Val/Met individuals who presented also the *DRD2* A1 allele, demonstrating also a gene-gene interaction.

Luijk et al. (2011) investigated the association of several genetic polymorphisms and infant attachment security and disorganization in a sample composed of over 500 children from two different cohorts. *COMT* Val158Met heterozygotes were more disorganized in both samples (combined effect size $d = 0.22$, $CI95 = 0.10-0.34$, $p < 0.001$), which the authors considered an example of negative heterosis.

Costas et al. (2011) investigated the hypothesis of overdominance (a. k. a., heterosis) in two samples of persons having schizophrenia ($n = 762$) and controls ($n = 1,042$). In these samples, they detected a protective effect against schizophrenia of the *COMT* Val/Met heterozygous genotype ($OR = 0.75$, $CI95 = 0.62-0.91$, $p = 0.003$). In addition, they conducted a meta-analysis including 13,894 schizophrenic patients and 16,087 controls from 51 studies. A protective effect of the Val/Met genotype was also detected (pooled $OR = 0.946$, $CI95 = 0.904-0.989$, $p = 0.015$).

It is important to consider that heterosis is by far the less investigated hypothesis regarding *COMT* effects on behavior. The wild-type allele at the 158 position is a valine. The mutation Val158Met is an evolutionary novelty present in the human, but not in the gorilla, chimpanzee, bonobo, and orangutan (Piffer, 2013). Currently, the frequency of

the Met allele is usually high (20–60%; Piffer, 2013) in most of the populations reported so far. This is surprising, considering that the enzyme activity is importantly reduced by the Met allele. The high frequency of the Met allele suggests that some selection mechanism is in place. From the literature review presented here, two main possible mechanisms emerge. The Met allele may have reached high frequencies because Met/- genotypes are advantageous for some *COMT* related phenotypes. Alternatively, heterosis itself is advantageous because intermediate dopamine levels at the synaptic cleft would be more adaptive, under usual environmental conditions, than high or low levels (Arnsten, 1998). The same could happen in the case of MA. Our data suggest a positive heterosis model. Heterozygous individuals exhibit MA levels closer to the grand average, and are less susceptible to worries related to math performance. Males having both homozygous phenotypes present lower MA than all other groups. Females having both homozygous phenotypes present higher levels of MA than all other groups.

It is important to note some limitations in our study. First, the specificity of the MA construct could not be investigated, as we did not include measures of achievement in other domains (such as reading or spelling), as well as other anxiety-related constructs such as self-efficacy and attitudes toward school performance in general and generalized anxiety. Second, the sample size is considerable, but still not enough for an analysis of different genotype groups separately, particularly when considering only the boys or only the girls with the Met/Met genotype. Third, MA is not caused by a single gene, so that many more candidate genes and environmental factors will need to be studied. Fourth, MA probably results from a combination of math-specific factors and general anxiety. The present study has ruled out the likelihood of this gene operating by affecting math ability, but it is not clear as yet whether it could be operating by affecting general anxiety.

Notwithstanding its limitations, the present study adds important information to the knowledge of the neurogenetic underpinnings of MA: (a) a thorough understanding of the origins of MA requires considerations of both environmental and genetic factors; (b) the dopaminergic system, a multifunctional system especially important in human evolution (Previc, 2009; Piffer, 2013), is also relevant for clarifying the neurobiological underpinnings of MA; (c) testing for associations between psychological phenotypes and single-loci genetic markers should consider all possible genetic models (dominance, codominance, and heterosis); and (d) sex differences in MA associated with the *COMT* Val158Met polymorphism are detectable even before puberty. Sex differences in the effect of Val158Met polymorphism in prepubertal children have already been described for cognitive functions (Barnett et al., 2007). Future research should investigate whether the heterosis model of the interaction among sex, *COMT* Val158Met polymorphism and MA is generalizable to other forms of anxiety. An epigenetic research approach is required to address interactions among the *COMT* Val158Met polymorphism with other dopaminergic and non-dopaminergic genes and with sex-hormonal and other metabolic pathways.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Resolução no. 196/96 and Resolução 466/12, of the Conselho Nacional de Saúde of the Brazilian Ministério da Saúde. The protocols were approved by the Comitê de Ética em Pesquisa da Universidade Federal de Minas Gerais. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

MRSC and VH designed the experiments, supervised data collection and analyses, and wrote the manuscript. GW analyzed results and wrote the manuscript. AM, AJ-C, MA, and MM collected and analyzed data and helped to write the manuscript.

FUNDING

This study was supported by grants from the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, APQ-02755-SHA, APQ-03289-10, APQ-02953-14, and APQ-03642-12). VH was supported by a CNPq fellowship (409624/2006-3, 308157/2011-7, and 308267/2014-1)

REFERENCES

- Angelini, A. L., Alves, I. C. B., Custódio, E. M., Duarte, W. F., and Duarte, J. L. M. (1999). *Matrizes Progressivas Coloridas de Raven – Escala Especial*. São Paulo: Centro Editor de Testes e Pesquisas em Psicologia.
- Arnsten, A. F. T. (1998). Catecholamine modulation of prefrontal cognitive function. *Trends Cogn. Sci.* 2, 436–447. doi: 10.1016/s1364-6613(98)01240-6
- Ashcraft, M. H., Krause, J. A., and Hopko, D. R. (2007). “Is math anxiety a mathematical learning disability?” in *Why is Math so Hard for some Children? The Nature and Origins of Mathematical Learning Difficulties and Disabilities*, eds D. B. Berch and M. M. M. Mazzocco (Baltimore, MD: Brookes), 329–348. doi: 10.1097/DBP.0b013e31817aef8
- Ashton, K. A., Meldrum, C. J., McPhillips, M. L., Suchy, J., Kurzwaski, G., Lubinski, J., et al. (2006). The association of the COMT V158M polymorphism with endometrial/ovarian cancer in HNPCC families adhering to the Amsterdam criteria. *Hered. Cancer Clin. Pract.* 15, 94–102. doi: 10.1186/1897-4287-4-2-94
- Associação Brasileira de Empresas de Pesquisa [ABEP] (2018). *Crítico de Classificação Econômica Brasil*.
- Barnett, J. H., Heron, J., Ring, S. M., Golding, J., Goldman, D., Xu, K., et al. (2007). Gender-specific effects of the catechol-O-methyltransferase val 108/158 met polymorphism on cognitive function in children. *Am. J. Psychiatry* 164, 142–149. doi: 10.1176/ajp.2007.164.1.142
- Barzman, D., Geise, C., and Lin, P. I. (2015). Review of the genetic basis of emotion dysregulation in children and adolescents. *World J. Psychiatry* 5, 112–117. doi: 10.5498/wjp.v5.i1.112
- Baumann, C., Klauke, B., Weber, H., Domschke, K., Zwanzer, P., Pauli, P., et al. (2013). The interaction of early life experiences with COMT val158met affects anxiety sensitivity. *Genes Brain Behav.* 12, 821–829. doi: 10.1111/gbb.12090
- Beilock, S. L., Gunderson, E. A., Ramirez, G., and Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1860–1863. doi: 10.1073/pnas.0910967107
- Bonvicini, C., Faraone, S. V., and Scassellati, C. (2016). Attention-deficit hyperactivity disorder in adults: a systematic review and meta-analysis of genetic, pharmacogenetic and biochemical studies. *Mol. Psychiatry* 21, 872–884. doi: 10.1038/mp.2016.74

and Programa de Capacitação em Neuropsicologia do Desenvolvimento (FEAPAEs-MG, APAE-BH, PRONAS-Ministério da Saúde, Brazil). VH participates in the INCT-ECCE, which was supported by the following grants: FAPESP: 2014/50909-8, CNPQ: 465686/2014-1, and CAPES: 88887.136407/2017-00. MRSC was supported by a CNPq fellowship (312068/2015-8). GW was supported by a grant from the University of Graz (Unkonventionelle Forschung, nr. AVO160200008). AJ-C and AM were supported by fellowships from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

ACKNOWLEDGMENTS

We thank the children, their parents, and also the principals of the schools for taking part in this research. We thank Mr. Peter Laspina, from ViaMundi Idiomas e Traduções for reviewing this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01013/full#supplementary-material>

- Brosnan, M., Gallop, V., Iftikhar, N., and Keogh, E. (2011). Digit ratio (2D:4D), academic performance in computer science and computer-related anxiety. *Pers. Individ. Dif.* 51, 371–375. doi: 10.1016/j.paid.2010.07.009
- Bull, R., Davidson, W. A., and Nordmann, E. (2010). Prenatal testosterone, visual-spatial memory, and numerical skills in young children. *Learn. Individ. Dif.* 20, 246–250. doi: 10.1016/j.lindif.2009.12.002
- Burnham, K. P., and Anderson, D. R. (2003). *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach*. Berlin: Springer Science+Business Media.
- Campbell, L. E., Azuma, R., Ambery, F., Stevens, A., Smith, A., Morris, R. G., et al. (2010). Executive functions and memory abilities in children with 22q11.2 deletion syndrome. *Aust. N. Z. J. Psychiatry* 44, 364–371. doi: 10.3109/00048670903489882
- Carmel, M., Zarchi, O., Michaelovsky, E., Frisch, A., Patya, M., Green, T., et al. (2014). Association of COMT and PRODH gene variants with intelligence quotient (IQ) and executive functions in 22q11.2DS subjects. *J. Psychiatr. Res.* 56, 28–35. doi: 10.1016/j.jpsychires.2014.04.019
- Ceci, S. J., Ginther, D. K., Kahn, S., and Williams, W. M. (2014). Women in academic science? a changing landscape. *Psychol. Sci. Public Interest* 15, 75–141. doi: 10.1177/1529100614541236
- Chapman, B. P., Duberstein, P. R., Sörensen, S., and Lyness, J. M. (2007). Gender differences in Five Factor Model personality traits in an elderly cohort. *Pers. Individ. Dif.* 43, 1594–1603. doi: 10.1016/j.paid.2007.04.028
- Charlesworth, D., and Willis, J. H. (2009). The genetics of inbreeding depression. *Nat. Rev. Genet.* 10, 783–796. doi: 10.1038/nrg2664
- Chen, C., Chen, C., Moyzis, R., Dong, Q., He, Q., Zhu, B., et al. (2011). Sex modulates the associations between the COMT gene and personality traits. *Neuropsychopharmacology* 36, 1593–1598. doi: 10.1038/npp.2011.39
- Chen, J., Lipska, B. K., Halim, N., Ma, Q. D., Matsumoto, M., Melhem, S., et al. (2004). Functional analysis of genetic variation in catechol-O-methyltransferase (COMT): effects on mRNA, protein, and enzyme activity in postmortem human brain. *Am. J. Hum. Genet.* 75, 807–821. doi: 10.1086/425589
- Comings, D. E., and MacMurray, J. P. (2000). Molecular heterosis: a review. *Mol. Genet. Metab.* 71, 19–31. doi: 10.1006/mgme.2000.3015

- Cordes, S., Williams, C. L., and Meck, W. H. (2007). Common representations of abstract quantities. *Curr. Dir. Psychol. Sci.* 16, 156–161. doi: 10.1111/j.1467-8721.2007.00495.x
- Costa, A. J., Silva, J. B. L., Chagas, P. P., Krinzinger, H., Lonneman, J., Willmes, K., et al. (2011). A hand full of numbers: a role for offloading in arithmetics learning? *Front. Psychol.* 2:368. doi: 10.3389/fpsyg.2011.00368
- Costas, J., Sanjuán, J., Ramos-Ríos, R., Paz, E., Agra, S., Ivorra, J. L., et al. (2011). Heterozygosity at catechol-O-methyltransferase Val158Met and schizophrenia: new data and meta-analysis. *J. Psychiatr. Res.* 45, 7–14. doi: 10.1016/j.jpsychires.2010.04.021
- Coull, J. T., Cheng, R. K., and Meck, W. H. (2011). Neuroanatomical and neurochemical substrates of timing. *Neuropsychopharmacology* 36, 3–25. doi: 10.1038/npp.2010.113
- Dawling, S., Roodi, N., Mernaugh, R. L., Wang, X., and Parl, F. F. (2001). Catechol-methyltransferase (COMT)-mediated metabolism of catechol estrogens: comparison of wild-type and variant COMT isoforms. *Cancer Res.* 61, 6716–6722.
- de Bruin, E. I., Verheij, F., Wiegman, T., and Ferdinand, R. F. (2006). Differences in finger length ratio between males with autism, pervasive developmental disorder-not otherwise specified, ADHD, and anxiety disorders. *Dev. Med. Child Neurol.* 48, 962–965. doi: 10.1111/j.1469-8749.2006.tb01266.x
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-n
- Dickinson, D., and Elvevag, B. (2009). Genes, cognition and brain through a COMT lens. *Neuroscience* 164, 72–87. doi: 10.1016/j.neuroscience.2009.05.014
- Domschke, K., Baune, B. T., Havlik, L., Stuhmann, A., Suslow, T., Kugel, H., et al. (2012). Catechol-O-methyltransferase gene variation: impact on amygdala response to aversive stimuli. *Neuroimage* 60, 2222–2229. doi: 10.1016/j.neuroimage.2012.02.039
- Domschke, K., Freitag, C. M., Kühlenbäumer, G., Schirmacher, A., Sand, P., Nyhuis, P., et al. (2004). Association of the functional V158M catechol-O-methyltransferase polymorphism with panic disorder in women. *Int. J. Neuropsychopharmacol.* 7, 183–188. doi: 10.1017/S146114570400416X
- Dowker, A., Bennett, K., and Smith, L. (2012). Attitudes to mathematics in primary school children. *Child Dev. Res.* 2012:124939. doi: 10.1155/2012/124939
- Dowker, A., Sarkar, A., and Looi, C. Y. (2016). Mathematics anxiety: what have we learned in 60 years? *Front. Psychol.* 7:508. doi: 10.3389/fpsyg.2016.00508
- Eccles, J. S., Jacobs, J. E., and Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *J. Soc. Issues* 46, 183–201. doi: 10.1111/j.1540-4560.1990.tb01929.x
- Enoch, M. A., Waheed, J. F., Harris, C. R., Albaugh, B., and Goldman, D. (2009). COMT Val158Met and cognition: main effects and interaction with educational attainment. *Genes Brain Behav.* 8, 36–42. doi: 10.1111/j.1601-183X.2008.00441.x
- Enoch, M. A., Xu, K., Ferro, E., Harris, C. R., and Goldman, D. A. (2003). Genetic origins of anxiety in women: a role for a functional catechol-o-methyltransferase polymorphism. *Psychiatr. Genet.* 13, 33–41. doi: 10.1097/01.ypg.0000054709.85338.c3
- Evans, J., Xu, K., Heron, J., Enoch, M. A., Araya, R., Lewis, G., et al. (2009). Emotional symptoms in children: the effect of maternal depression, life events, and COMT genotype. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 150, 209–218. doi: 10.1002/ajmg.b.30789
- Ferreira, F. D. O., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., et al. (2012). Explaining school mathematics performance from symbolic and nonsymbolic magnitude processing: similarities and differences between typical and low-achieving children. *Psychol. Neurosci.* 5, 37–46. doi: 10.3922/j.psns.2012.1.06
- Fink, B., Brookes, H., Neave, N., Manning, J. T., and Geary, D. D. (2006). Second to fourth digit ratio and numerical competence in children. *Brain Cogn.* 61, 211–218. doi: 10.1016/j.bandc.2006.01.001
- Franconi, C. P., McDonald-McGinn, D., Zackai, E. H., McNamara, M. A., Salmons, H., Moss, E., et al. (2016). IQ and hemizygosity for the Val(158) Met functional polymorphism of COMT in 22q11DS. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 171, 1112–1115. doi: 10.1002/ajmg.b.32492
- Gauderman, W. J. (2002). Sample size requirements for association studies of gene-gene interaction. *Am. J. Epidemiol.* 155, 478–484. doi: 10.1093/aje/155.5.478
- Gauderman, W. J. (2003). Candidate gene association analysis for a quantitative trait, using parent-offspring trios. *Genet. Epidemiol.* 25, 327–338. doi: 10.1002/gepi.10262
- Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., et al. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS One* 8:e67374. doi: 10.1371/journal.pone.0067374
- Gogos, J. A., Morgan, M., Luine, V., Santha, M., Ogawa, S., Pfaff, D., et al. (1998). Catechol-O-methyltransferase-deficient mice exhibit sexually dimorphic changes in catecholamine levels and behavior. *Proc. Natl. Acad. Sci. U.S.A.* 95, 9991–9996. doi: 10.1073/pnas.95.17.9991
- Goldberg, T. E., Egan, M. F., Gscheidle, T., Coppola, R., Weickert, T., Kolachana, B. S., et al. (2003). Executive subprocesses in working memory: relationship to catechol-O-methyltransferase Val158Met genotype and schizophrenia. *Arch. Gen. Psychiatry* 60, 889–896.
- González-Castro, T. B., Hernández-Díaz, Y., Juárez-Rojop, I. E., López-Narváez, M. L., Tovilla-Zárate, C. A., and Fresan, A. (2016). The role of a catechol-O-methyltransferase (COMT) Val158Met genetic polymorphism in schizophrenia: a systematic review and updated meta-analysis on 32,816 subjects. *Neuromolecular Med.* 18, 216–231. doi: 10.1007/s12017-016-8392-z
- Gosso, M. F., Geus, E. J., Polderman, T. J., Boomsma, D. I., Heutink, P., and Posthuma, D. (2008). Catechol O-methyl transferase and dopamine D2 receptor gene polymorphisms: evidence of positive heterosis and gene-gene interaction on working. *Eur. J. Hum. Genet.* 16, 1075–1082. doi: 10.1038/ejhg.2008.57
- Gottschalk, M. G., and Domschke, K. (2017). Genetics of generalized anxiety disorder and related traits. *Dialogues Clin. Neurosci.* 19, 159–168.
- Gunderson, E. A., Ramirez, G., Levine, S. C., and Beilock, S. L. (2012). The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles* 66, 153–166. doi: 10.1007/s11199-011-9996-2
- Haase, V. G., Guimarães, A. P. L., and Wood, G. (2019). “Math & emotions: the case of math anxiety,” in *International Handbook of Math Learning Difficulties: from the Lab to the Classroom*, eds A. Fritz, V. G. Haase, and P. Räsänen (São Paulo: Springer), 469–503.
- Haase, V. G., Júlio-Costa, A., Pinheiro-Chagas, P., Oliveira, L. D. F. S., Micheli, L. R., and Wood, G. (2012). Math self-assessment, but not negative feelings, predicts mathematics performance of elementary school children. *Child Dev. Res.* 2012:982672. doi: 10.1155/2012/982672
- Harrison, P. J., and Tunbridge, E. M. (2008). Catechol-O-methyltransferase (COMT): a gene contributing to sex differences in brain function, and to sexual dimorphism in the predisposition to psychiatric disorders. *Neuropsychopharmacology* 33, 3037–3045. doi: 10.1038/sj.npp.1301543
- Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *J. Res. Math. Educ.* 21, 33–46.
- Hosák, L. (2007). Role of the COMT gene Val158Met polymorphism in mental disorders: a review. *Eur. Psychiatry* 22, 276–281. doi: 10.1016/j.eurpsy.2007.02.002
- Ireson, G. (2016). Gender achievement and social, political and economic equality: a European perspective. *Educ. Stud.* 43, 40–50. doi: 10.1080/03055698.2016.1237868
- Jacobs, E., and D'Esposito, M. (2011). Estrogen shapes dopamine-dependent cognitive processes: implications for women's health. *J. Neurosci.* 31, 5286–5293. doi: 10.1523/JNEUROSCI.6394-10.2011
- Jiang, H., Xie, T., Ramsden, D. B., and Ho, S. L. (2003). Human catechol-O-methyltransferase down-regulation by estradiol. *Neuropharmacology* 45, 1011–1018. doi: 10.1016/s0028-3908(03)00286-7
- Júlio-Costa, A., Antunes, A. M., de Almeida Prado, A. C., Carvalho, M. R. S., and Haase, V. G. (2014). Association between COMT val158met polymorphism and working memory tasks in children and adolescents: a systematic review. *Neurotransmitter* 2:e483. doi: 10.14800/nt.483
- Júlio-Costa, A., Antunes, A. M., Lopes-Silva, J. B., Moreira, B. C., Vianna, G. S., Wood, G., et al. (2013). Count on dopamine: influences of COMT polymorphisms on numerical cognition. *Front. Psychol.* 4:531. doi: 10.3389/fpsyg.2013.00531
- Karayorgou, M., Altemus, M., Galke, B. L., Goldman, D., and Murphy, D. L. (1997). Genotype determining low catechol-O-methyltransferase activity as a risk factor for obsessive-compulsive disorder. *Proc. Natl. Acad. Sci. U.S.A.* 94, 4572–4575. doi: 10.1073/pnas.94.9.4572

- Karayorgou, M., Simon, T. J., and Gogos, J. A. (2010). 2q11.2 microdeletions: linking DNA structural variation to brain dysfunction and schizophrenia. *Nat. Rev. Neurosci.* 11, 402–416. doi: 10.1038/nrn2841
- Kebir, O., and Joob, R. (2011). Neuropsychological endophenotypes in attention-deficit/hyperactivity disorder: a review of genetic association studies. *Eur. Arch. Psychiatry Clin. Neurosci.* 261, 583–594. doi: 10.1007/s00406-011-0207-5
- Krendl, A. C., Richeson, J. A., Kelley, W. M., and Heatherton, T. F. (2008). The negative consequences of threat A functional magnetic resonance imaging investigation of the neural mechanisms underlying women's underperformance in math. *Psychol. Sci.* 19, 168–175. doi: 10.1111/j.1467-9280.2008.02063.x
- Krinzinger, H., Kaufmann, L., Dowker, A., Thomas, G., Graf, M., Nuerk, H.-C., et al. (2007). Deutschsprachige version des fragebogens für rechenangst (FRA) für 6- bis 9-jährige Kinder. *Z. Kinder Jugendpsychiatr. Psychother.* 35, 341–351. doi: 10.1024/1422-4917.35.5.341
- Lee, L. O., and Prescott, C. A. (2014). Association of the catechol-O-methyltransferase val158met polymorphism and anxiety-related traits: a meta-analysis. *Psychiatr. Genet.* 24, 52–69. doi: 10.1097/YPG.000000000000018
- Lehto, K., Akkermann, K., Parik, J., Veidebaum, T., and Harro, J. (2013). Effect of COMT Val158Met polymorphism on personality traits and educational attainment in a longitudinal population representative study. *Eur. Psychiatry* 28, 492–498. doi: 10.1016/j.eurpsy.2013.06.008
- Leslie, A. M., Gallistel, C. R., and Gelman, R. (2007). “Where do integers come from,” in *The Innate Mind: Foundations and the Future*, Vol. 3, eds P. Carruthers, S. Laurence, and S. Stich (New York, NY: Oxford University Press), 109–149.
- Luijck, M. P., Roisman, G. I., Haltigan, J. D., Tiemeier, H., Booth-Laforce, C., van Ijzendoorn, M. H., et al. (2011). Dopaminergic, serotonergic, and oxytonergic candidate genes associated with infant attachment security and disorganization? In search of main and interaction effects. *J. Child Psychol. Psychiatry* 52, 1295–1307. doi: 10.1111/j.1469-7610.2011.02440.x
- Malanchini, M., Rimpf, K., Shakeshaft, N. G., Rodic, M., Schofield, K., Selzam, S., et al. (2017). The genetic and environmental aetiology of spatial, mathematics and general anxiety. *Sci. Rep.* 7:42218. doi: 10.1038/srep42218
- Maloney, E. A., Waechter, S., Risko, E. F., and Fugelsang, J. A. (2012). Reducing the sex difference in math anxiety: the role of spatial processing ability. *Learn. Individ. Differ.* 22, 380–384. doi: 10.1016/j.lindif.2012.01.001
- Mandelbaum, E. (2013). Numerical architecture. *Top. Cogn. Sci.* 5, 367–386. doi: 10.1111/tops.12014
- Mata, M. D. L., Monteiro, V., and Peixoto, F. (2012). Attitudes towards mathematics: effects of individual, motivational, and social support factors. *Child Dev. Res.* 2012:876028. doi: 10.1155/2012/876028
- Mathew, S. J., and Ho, S. (2006). Etiology and neurobiology of social anxiety disorder. *J. Clin. Psychiatry* 67, 9–13.
- McLean, C. P., Asnaani, A., Litz, B. T., and Hofmann, S. G. (2011). Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. *J. Psychiatr. Res.* 45, 1027–1035. doi: 10.1016/j.jpsychires.2011.03.006
- McMullan, M., Jones, R., and Lea, S. (2012). Math anxiety, self-efficacy, and ability in British undergraduate nursing students. *Res. Nurs. Health* 35, 178–186. doi: 10.1002/nur.21460
- Mejias, S., Mussolin, C., Rousselle, L., Grégoire, J., and Noël, M. P. (2012). Numerical and nonnumerical estimation in children with and without mathematical learning disabilities. *Child Neuropsychol.* 18, 550–575. doi: 10.1080/09297049.2011.625355
- Mier, D., Kirsch, P., and Meyer-Lindenberg, A. (2010). Neural substrates of pleiotropic action of genetic variation in COMT: a meta-analysis. *Mol. Psychiatry* 15, 918–927. doi: 10.1038/mp.2009.36
- Miller, G. A., and Chapman, J. P. (2001). Misunderstanding analysis of covariance. *J. Abnorm. Psychol.* 110, 40–48.
- Miller, S. A., Dykes, D. D., and Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16:1215. doi: 10.1093/nar/16.3.1215
- Minato, S., and Yanase, S. (1984). On the relationship between students attitudes towards school mathematics and their levels of intelligence. *Educ. Stud. Math.* 15, 313–320. doi: 10.1007/bf00312080
- Moenik, M., and Zahed-Babelan, A. (2010). A study of simple and multiple relations between mathematics attitude, academic motivation and intelligence quotient with mathematics achievement. *Procedia Soc. Behav. Sci.* 2, 1537–1542. doi: 10.1016/j.sbspro.2010.03.231
- Montag, C., Jurkiewicz, M., and Reuter, M. (2012). The role of the catechol-O-methyltransferase (COMT) gene in personality and related psychopathological disorders. *CNS Neurol. Disord. Drug Targets* 11, 236–250. doi: 10.2174/187152712800672382
- Nikolac Perkovic, M., Nedic Erjavec, G., Stefulj, J., Muck-Seler, D., Pivac, N., Kocijan Hercigonja, D., et al. (2014). Association between the polymorphisms of the selected genes encoding dopaminergic system with ADHD and autism. *Psychiatry Res.* 215, 260–261. doi: 10.1016/j.psychres.2013.10.018
- Oliveira-Ferreira, F., Costa, D., Micheli, L., Oliveira, L., Pinheiro-Chagas, P., and Haase, V. (2012). School Achievement Test: normative data for a representative sample of elementary school children. *Psychol. Neurosci.* 5, 157–164. doi: 10.3922/j.pns.2012.2.05
- Olsson, C. A., Anney, R. J. L., Lotfi-Miri, M., Byrnes, G. B., Williamson, R., and Patton, G. C. (2005). Association between the COMT Val158Met polymorphism and propensity to anxiety in an Australian population-based longitudinal study of adolescent health. *Psychiatr. Genet.* 15, 109–115. doi: 10.1097/00041444-200506000-00007
- Olsson, C. A., Byrnes, G. B., Anney, R. J. L., Collins, V., Hemphill, S. A., Williamson, R., et al. (2007). COMT Val158Met and 5HTTLPR functional loci interact to predict persistence of anxiety across adolescence: results from the Victorian Adolescent Health Cohort Study. *Genes Brain Behav.* 6, 647–652. doi: 10.1111/j.1601-183X.2007.00313.x
- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012
- Piffer, D. (2013). Correlation of the COMT Val158Met polymorphism with latitude and a hunter-gather lifestyle suggests culture–gene coevolution and selective pressure on cognition genes due to climate. *Anthropol. Sci.* 121, 161–171. doi: 10.1537/ase.130731
- Pind, J., Gunnarsdottir, E. K., and Johannesson, H. S. (2003). Raven's Standard Progressive Matrices: new school age norms and a study of the test's validity. *Pers. Individ. Dif.* 34, 375–386. doi: 10.1016/s0191-8869(02)00058-2
- Pinheiro-Chagas, P., Wood, G., Knops, A., Krininger, H., Lonnemann, J., Starling-Alves, I., et al. (2014). In how many ways is the approximate number system associated with exact calculation? *PLoS One* 9:e111155. doi: 10.1371/journal.pone.0111155
- Pooley, E. C., Fineberg, N., and Harrison, P. J. (2007). The met(158) allele of catechol-O-methyltransferase (COMT) is associated with obsessive-compulsive disorder in men: case-control study and meta-analysis. *Mol. Psychiatry* 12, 556–561. doi: 10.1038/sj.mp.4001951
- Previc, F. H. (2009). *The Dopaminergic mind in Human Evolution and History*. New York, NY: Cambridge University Press.
- Primi, R., Ferrão, M. E., and Almeida, L. S. (2010). Fluid intelligence as a predictor of learning a longitudinal multilevel approach applied to math. *Learn. Individ. Dif.* 20, 446–451. doi: 10.1016/j.lindif.2010.05.001
- Ramirez, G., Gunderson, E. A., Levine, S. C., and Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *J. Cogn. Dev.* 14, 187–202. doi: 10.1080/15248372.2012.664593
- Raymond, M., and Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86, 248–249. doi: 10.1093/oxfordjournals.jhered.a111573
- Riccardi, P., Park, S., Anderson, S., Doop, M., Ansari, M. S., Schmidt, D., et al. (2011). Sex differences in the relationship of regional dopamine release to affect and cognitive function in striatal and extra striatal regions using positron emission tomography and fallypride. *Synapse* 65, 99–102. doi: 10.1002/syn.20822
- Riccardi, P., Zald, D., Li, R., Park, S., Ansari, M. S., Dawant, B., et al. (2006). Sex differences in amphetamine-induced displacement of [(18)F]fallypride in striatal and extrastriatal regions: a PET study. *Am. J. Psychiatry* 163, 1639–1641. doi: 10.1176/ajp.2006.163.9.1639
- Rohde, T. E., and Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence* 35, 83–92. doi: 10.1016/j.intell.2006.05.004
- Rothe, C., Koszycki, D., Bradwejn, J., King, N., Deluca, V., Tharmalingam, S., et al. (2006). Association of the Val158Met catechol O-methyltransferase genetic

- polymorphism with panic disorder. *Neuropsychopharmacology* 31, 2237–2242. doi: 10.1038/sj.npp.1301048
- Rousset, F. (2008). Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Mol. Ecol. Resour.* 8, 103–106. doi: 10.1111/j.1471-8286.2007.01931.x
- Ruiz-Sanz, J. I., Aurrekoetxea, I., Ruiz Del Agua, A., and Ruiz-Larrea, M. B. (2007). Detection of catechol-O-methyltransferase Val158Met polymorphism by a simple one-step tetra-primer amplification refractory mutation system-PCR. *Mol. Cell. Probes* 21, 202–207. doi: 10.1016/j.mcp.2006.12.001
- Shashi, V., Howard, T. D., Keshavan, M. S., Kaczorowski, J., Berry, M. N., Schoch, K., et al. (2010). COMT and anxiety and cognition in children with chromosome 22q11.2 deletion syndrome. *Psychiatry Res.* 178, 433–436. doi: 10.1016/j.psychres.2010.04.048
- Shashi, V., Keshavan, M. S., Howard, T. D., Berry, M. N., Basehore, M. J., Lewandowski, E., et al. (2006). Cognitive correlates of a functional COMT polymorphism in children with 22q11.2 deletion syndrome. *Clin. Genet.* 69, 234–238. doi: 10.1111/j.1399-0004.2006.00569.x
- Sheikh, H. I., Kryski, K. R., Kotelnikova, Y., Hayden, E. P., and Singh, S. M. (2017). Catechol-O-Methyltransferase gene (val158met) polymorphisms and anxious symptoms in early childhood: the roles of hypothalamus-pituitary-adrenal axis reactivity and life stress. *Neurosci. Lett.* 659, 86–91. doi: 10.1016/j.neulet.2017.08.060
- Sheikh, H. I., Kryski, K. R., Smith, H. J., Dougherty, L. R., Klein, D. N., Bufferd, S., et al. (2013). Catechol-O-methyltransferase gene val158met polymorphism and depressive symptoms during early childhood. *Am. J. Med. Genet. Part B* 162, 245–252. doi: 10.1002/ajmg.b.32141
- Sherwin, B. B. (2007). The clinical relevance of the relationship between estrogen and cognition in women. *J. Steroid Biochem. Mol. Biol.* 106, 151–156. doi: 10.1016/j.jsbmb.2007.05.016
- Siegler, R. S., and Braithwaite, D. W. (2017). Numerical development. *Annu. Rev. Psychol.* 68, 187–213. doi: 10.1146/annurev-psych-010416-044101
- Spencer, S. J., Steele, C. M., and Quinn, D. M. (1999). Stereotype threat and women's math performance. *J. Exp. Soc. Psychol.* 35, 4–28. doi: 10.1006/jesp.1998.1373
- Stein, D. J., Newman, T. K., Savitz, J., and Ramesar, R. (2006). Warriors versus worriers: the role of COMT gene variants. *CNS Spectr.* 11, 745–748. doi: 10.1017/s1092852900014863
- Stein, L. M. (1994). *TDE. Teste de Desempenho Escolar. Manual Para Aplicação e Interpretação*. São Paulo: Casa do Psicólogo.
- Stoet, G., and Geary, D. C. (2012). Can stereotype threat explain the gender gap in mathematics performance and achievement? *Am. Psychol. Assoc.* 16, 93–102. doi: 10.1037/a0026617
- Stoet, G., and Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: within-and across-nation assessment of 10 years of PISA data. *PLoS One* 8:e57988. doi: 10.1371/journal.pone.0057988
- Stoet, G., and Geary, D. C. (2015). Sex differences in academic achievement are not related to political, economic or social equality. *Intelligence* 48, 137–151. doi: 10.1073/pnas.1707719114
- Stoet, G., and Geary, D. C. (2016). Challenges for determining the causal effects between social behavior and testosterone. *Proc. Natl. Acad. Sci. U.S.A.* 113:E499.
- Szűcs, D., Nobes, A., Devine, A., Gabriel, F. C., and Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Front. Psychol.* 4:444. doi: 10.3389/fpsyg.2013.00444
- Tan, H. Y., Chen, Q., Goldberg, T. E., Mattay, V. S., Meyer-Lindenberg, A., Weinberger, D. R., et al. (2007). Catechol-O-methyltransferase Val158Met modulation of prefrontal-parietal-striatal brain systems during arithmetic and temporal transformations in working memory. *J. Neurosci.* 27, 13393–13401. doi: 10.1523/jneurosci.4041-07.2007
- Thomas, G., and Dowker, A. (2000). "Mathematics anxiety and related factors in young children," in *Paper Presented at the British Psychological Society Developmental Section Conference*, Bristol.
- Tosto, M. G., Malykh, S., Haworth, C. M. A., Thompson, L., Opfer, J., Petrill, S. A., et al. (2017). Number sense and mathematics: which, when, and how? *Dev. Psychol.* 53, 1924–1939. doi: 10.1037/dev0000331
- Tosto, M. G., Petrills, S. A., Halberda, J., Trzaskowski, M., Tikhomirova, T. N., Bogdanova, O. Y., et al. (2014). Why do we differ in number sense? Evidence from a genetically sensitive investigation. *Intell.* 43, 35–46. doi: 10.1016/j.intell.2013.12.007
- Tunbridge, E. M. (2010). "The catechol-O-methyltransferase gene: its regulation and polymorphisms," in *Basic Aspects of Catechol-O-Methyltransferase and the Clinical Applications of its Inhibitors*, ed. E. Nissinen (San Diego, CA: Elsevier Science Publishing Co Inc), 95–97.
- Utsumi, M. C., and Mendes, C. R. (2000). Researching the attitudes towards mathematics in basic education. *Educ. Psychol.* 20, 237–243. doi: 10.1080/713663712
- Wagenmakers, E. J., and Farrell, S. (2004). AIC model selection using Akaike weights. *Psychon. Bull. Rev.* 11, 192–196. doi: 10.3758/bf03206482
- Wahlstrom, D., White, T., and Luciana, M. (2010). Neurobehavioral evidence for changes in dopamine system activity during adolescence. *Neurosci. Biobehav. Rev.* 34, 631–648. doi: 10.1016/j.neubiorev.2009.12.007
- Wai, J., Cacchio, M., Putallaz, M., and Makel, M. C. (2010). Sex differences in the right tail of cognitive abilities: a 30 year examination. *Intelligence* 38, 412–423. doi: 10.1016/j.intell.2010.04.006
- Wang, Z., Hart, S. A., Kovas, Y., Lukowski, S., Soden, B., Thompson, L. A., et al. (2014). Who is afraid of math? Two sources of genetic variance for mathematical anxiety. *J. Child Psychol. Psychiatry* 55, 1056–1064. doi: 10.1111/jcpp.12224
- Wigfield, A., and Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *J. Educ. Psychol.* 80, 210–216. doi: 10.1037//0022-0663.80.2.210
- Woo, J. M., Yoon, K. S., Choi, Y. H., Oh, K. S., Lee, Y. S., and Yu, B. H. (2004). The association between panic disorder and the L/L genotype of catechol-O-methyltransferase. *J. Psychiatr. Res.* 38, 365–370. doi: 10.1016/j.jpsychires.2004.01.001
- Wood, G., Pinheiro-Chagas, P., Júlio-Costa, A., Micheli, L. R., Krinzinger, H., Kaufmann, L., et al. (2012). Math anxiety questionnaire: similar latent structure in Brazilian and German school children. *Child Dev. Res.* 2012:610192. doi: 10.1155/2012/610192
- Xie, T., Ho, S. L., and Ramsden, D. (1999). Characterization and implications of estrogenic down-regulation of human catechol-O-methyltransferase gene transcription. *Mol. Pharmacol.* 56, 31–38. doi: 10.1124/mol.56.1.31

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

Copyright © 2019 Júlio-Costa, Martins, Wood, Almeida, Miranda, Haase and Carvalho. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Assessing Mathematical School Readiness

Sandrine Mejias^{1*}, Claire Muller² and Christine Schiltz³

¹ CNRS, CHU Lille, UMR 9193 – SCALab – Sciences Cognitives et Sciences Affectives, Université de Lille, Lille, France,

² Luxembourg Centre for Educational Testing, Université du Luxembourg, Esch-sur-Alzette, Luxembourg, ³ Institute of Cognitive Science and Assessment, Université du Luxembourg, Esch-sur-Alzette, Luxembourg

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

Sabine Heim,
Rutgers, The State University
of New Jersey, United States
Jo Van Herwegen,
Kingston University, United Kingdom

*Correspondence:

Sandrine Mejias
sandrine.mejias@univ-lille.fr

Specialty section:

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

Received: 28 November 2018

Accepted: 03 May 2019

Published: 24 May 2019

Citation:

Mejias S, Muller C and Schiltz C
(2019) Assessing Mathematical
School Readiness.
Front. Psychol. 10:1173.
doi: 10.3389/fpsyg.2019.01173

Early math skills matter for later formal mathematical performances, academic and professional success. Accordingly, it is important to accurately assess mathematical school readiness (MSR) at the beginning of elementary school. This would help identifying children who are at risk of encountering difficulties in math and then stimulate their acquisition of mathematical skills as soon as possible. In the present study, we present a new test that allows professionals working with children (e.g., teachers, school psychologists, speech therapists, and school doctors) to assess children's MSR when they enter formal schooling in a simple, rapid and efficient manner. 346 children were assessed at the beginning of 1st Grade (6-to-7-year-olds) with a collective test assessing early mathematical abilities (T1). In addition, children's math skills were evaluated with classical curriculum math tests at T1 and a year later, in 2nd Grade (T2, 7-to-8-year-olds). After assessing internal consistency, three tasks were retained for the final version of the MSR test. Test performance confirmed to be essentially unidimensional and systematically related to the scores children obtained in classical tests in 1st and 2nd Grade. By using the present MSR test, it is possible to identify pupils at risk of developing low math skills right from the start of formal schooling in 1st Grade. Such a tool is needed, as children's level in math at school beginning (or school readiness) is known to be foundational for their future academic and professional carrier.

Keywords: mathematical school readiness, numeracy, math skills, number sense, arithmetic, mathematical learning

INTRODUCTION

Considering the importance of mathematics in modern society, math activities play a central role in a child's education. Building good math skills is an essential part of a first grader's learning process and determines academic success down the road. Indeed, it has been demonstrated that children's scholastic level at the beginning of formal schooling - or school readiness - is very important for their future academic and professional carriers (Currie and Thomas, 1999; Duncan et al., 2007; Romano et al., 2010). Especially, early math skills developed during kindergarten appear to be one of the most powerful predictors of later formal learning, including reading (Duncan et al., 2007; Pagani et al., 2010; Romano et al., 2010). In addition, many longitudinal studies have emphasized the importance of early math skills for the development of more elaborate mathematical abilities (e.g., Jordan et al., 2007, 2009; Aunio and Niemivirta, 2010). Moreover, young adults' proficiency to use simple math to solve problems encountered in everyday life seems to determine their likelihood of full-time employment (e.g., Rivera-Batiz, 1992).

Consequently, it seems highly relevant to evaluate early math skills of children at the beginning of formal (mathematical) learning, or in other words, their mathematical school readiness (MSR). This allows us to identify children with low math skills at the beginning of 1st Grade and to subsequently set up appropriate educational support measures for the children in need (National Research Council, 2001). These support measures will provide an ideal basis for later mathematical learning and prevent a vicious circle of poor basic skills leading to poor mathematical learning, which in turn results in numerical shortcomings. However, to be able to identify children in (great) difficulty, teachers need validated and standardized tools. Yet, they deplore the lack of such tools and indicate that they are usually forced to rely on their own “home-made” tools or intuition, which is not ideal and leaves them feeling uncomfortable (Do, 2007). Data showing that teachers’ judgments are perceptually biased and that they have difficulties judging their students’ cognitive potential confirm the actual problem of the situation (e.g., Fischbach et al., 2013).

Mathematical school readiness focuses on the narrow window of math development between the acquisition of (pre)mathematical precursor skills in kindergarten and the implementation of formal mathematical education in elementary school. The acquisition of the Arabic number notation system constitutes a key element of MSR, because it bridges the innate core magnitude system (e.g., Feigenson et al., 2004) and the development of the exact number representations underlying the (ordinal) mental number line and arithmetic thinking (see von Aster, 2000; von Aster and Shalev, 2007). According to von Aster and Shalev’s (2007) four-step-developmental model of number acquisition, the acquisition of the Arabic number system (i.e., the visual Arabic code, see also Dehaene, 1992) is a major challenge in the development of children’s math skills. This acquisition implies the progressive learning of visual number symbols (i.e., Arabic numbers), the place value syntax and the corresponding transcoding rules (see Thevenot and Fayol, 2018, for a review). Together with the verbal number system, which develops during preschool years, the acquisition of the Arabic notation system (including multi-digit numbers) implicitly starts in preschool (Gilmore et al., 2007; Mejias and Schiltz, 2013), probably because of the widespread use of digital displays in children’s direct environment. It is then systematically consolidated and enhanced during 1st Grade through formal and explicit instruction (Mix et al., 2014).

However, poorly developed mathematical competencies are observed in a non-negligible number of (young) children and adults (3–7%; Gross-Tsur et al., 1996; Shalev et al., 2005; Reigosa-Crespo et al., 2012; American Psychiatric Association [APA], 2013). According to the DSM-5 (2013), specific learning disorder is now a single, overall diagnosis, incorporating deficits that impact academic achievement. Specific learning disorder refers to significant and persistent difficulties in learning and using one’s cultural symbol systems (e.g., alphabet, characters, and Arabic numbers) that are required for skilled reading, writing, and math, and must be learned by instruction. Persons with specific learning disorder are unable to perform academically at a level appropriate to their intelligence and age. The definition states that difficulties should have persisted for at least 6 months

despite interventions, and skills should be substantially below those expected for that given age. It is now recommended to give this diagnosis only from 1.5 standard deviations below the mean for age (which correspond to a performance within the lowest 7% in standardized mathematical tests, while previously the lowest 10% was commonly accepted). Beside those children with *specific learning disorder in math*, individuals achieving between 11% and 25% in standardized mathematical tests are classically identified as *low math achievers* (see Geary, 2011, for a review).

Low math achievers and especially individuals with specific learning disorder in math already perform less accurately than typically developing children in 1st Grade (Geary, 2011). Moreover, those children who perform in the lowest quartile in curriculum math tests also experience difficulties in basic math skills. This includes the processing of numbers and numerosities (Koontz and Berch, 1996; Landerl et al., 2004; Rubinsten and Henik, 2005; Mussolin et al., 2010), even when the task simply requires to count small sets of 1 to 4 items (Willburger et al., 2008). Furthermore, there are small but systematic group differences between 1st Grade children with specific learning disorder in math and controls in number naming, number writing (Geary et al., 1999) and comparing the magnitude of one- or two-digit numbers (Landerl et al., 2004, 2009; Rousselle and Noël, 2007; Iuculano et al., 2008; Landerl and Kölle, 2009). In sum, low math achievers and children with specific learning disorder in math reveal atypical performances in tasks requiring identification, representation and production of numerical quantities and symbols, in number comparison, in counting as well as in (simple) math problems.

There are currently a number of curriculum math tests for young children, which aim to provide an exhaustive diagnoses of mathematical learning disabilities (e.g., Van Nieuwenhoven et al., 2008; Lafay and Helloin, 2016) or to rapidly screen children’s ability in magnitude comparison (Nosworthy et al., 2013; Brankaer et al., 2017), which is one of the major precursors of math skills (Halberda et al., 2008). The former offer very complete and detailed insights into a child’s mathematical ability, but they have to be administered by specifically qualified professionals (i.e., psychologists or speech therapists) in time-consuming individual testing sessions. The latter can be easily and quickly run in group settings by a wide range of school professionals, but they focus on a specific mathematical precursor ability and also require a specific psychological knowledge basis for interpreting the results and translating them into classroom practice. In contrast, there are currently no tests that allow school teachers to evaluate children’s early mathematical abilities, by administering and interpreting a validated and standardized test in the classroom setting. This is especially relevant and desirable at the beginning of formal schooling, because it allows teachers to identify those children with insufficient math skills directly at the beginning of the formal learning trajectory. Accordingly, teachers will be able to (a) set up appropriate learning and catch-up measures and/or (b) orient children toward special care. In summary, to the best of our knowledge, there are currently no tests that allow teachers to assess MSR based on psychometrically validated tasks with a high face-validity that can be easily administered in classroom settings.

Here we propose a test of MSR systematically assessing the mastery of visual number symbols at the entrance of formal schooling (i.e., at the beginning of 1st Grade). By this means, we intend to provide a psychometrically validated tool that can be easily used in classroom-settings and interpreted by school teachers. The MSR test therefore consists of different tasks having a high face-validity in the context of math education, while being also firmly embedded in neuro-psychological theories of typical and atypical numerical development. The test is composed of tasks probing Arabic number identification, writing Arabic numbers to dictation, writing Arabic numbers as a result of counting, Arabic number comparison, as well as basic arithmetic problem solving.

The present study aimed to evaluate the psychometric validity of the MSR test and its constituent tasks. Moreover, it determined the concurrent and predictive criterion validity of the test by evaluating whether 1st Graders' performances on the test could significantly predict their performances on formal mathematics tasks, evaluated at the time of testing and 1 year later in 2nd Grade of elementary school. If the test items are valid and allow predicting children's mathematical performance in 2nd Grade, then our test can help school teachers identify those children with insufficient MSR, thereby providing them with an empirical basis to orient these children toward dedicated educational support and special care measures.

MATERIALS AND METHODS

Participants

Totally 346 participants (163 boys) were included in the study. The mean age was 6.30 years [± 0.35].

Participants were recruited from twelve different public schools in Belgium, at the beginning of 1st Grade.

This study was carried out in accordance with the recommendations of the research ethics committee of the Université Catholique de Louvain (Belgium). The protocol was approved by the research ethics committee of the Université Catholique de Louvain (Belgium). Written informed parental consent was obtained for each of the children, in accordance with the Declaration of Helsinki.

The schools' socio-economic index level ranged from 4 to 20.¹ Participating schools were distributed in five different socio-economic index levels: One school was classified at very low level "4" (including 12 participants); two schools at intermediate level "12" (including 61 participants); four

schools at intermediate level "13" (including 93 participants); four schools at very high level "19" (including 166 participants), and one school at highest level "20" (including 14 participants). In each school, children were tested for the first time (T1) in mid-September (6-to-7-year-olds) at the beginning of 1st Grade and for the second time (T2) in mid-September 1 year later (7-to-8-year-olds) at the beginning of 2nd Grade. Children's age in months was similar across the five different socio-economic groups (with the largest age-difference in terms of months between two groups belonging to socio-economic index 12 and 13, $p = 0.08$; all other contrasts, $p > 0.3$). Data were collected by only one person.

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

Materials and Procedure

Mathematical School Readiness Test

To assess children's MSR when they enter formal schooling a collective test of early mathematical abilities was developed. Considering the neuro-cognitive literature on typical and atypical numerical development (e.g., von Aster and Shalev, 2007; Dowker, 2008; Geary, 2011) this test aims to describe children's abilities focusing on the mastery of visual number symbols typically required at the moment of formal (math) schooling entrance (i.e., at T1, during the first month of the 1st Grade). The test was designed to have a high face-validity for teachers and therefore includes all early math abilities described also in the school competence standards in Wallonia in Belgium (Van Lint, 2010) (i.e., visual number symbol identification, writing numbers to dictation, symbolic quantity representation, counting abilities and arithmetic abilities). The time required to complete the entire test was approximately 20 min. Five tasks were administered and evaluated (**Appendix A**):

- (1) *Identifying visual number symbols (number identification)*: Arabic digits (3, 6, 8, and 9) were presented on a sheet of paper amongst non-numerical stimuli (a, @, \$, and f). Children were asked to circle Arabic numbers ($N = 4$) and to cross out if not a number ($N = 4$). Note that if the child clearly identifies the numbers by circling them and does not cross out the other symbols, a maximum score of 8 is given. For example, a child who circles all the numbers and another non-numerical symbol is given a score of 7, etc. Average time required to achieve the subtest was 90 s.
- (2) *Writing numbers to dictation (number writing)*: Children were told to write down the number they hear in the correct box (boxes were presented sequentially in line, with each box having a different color to prevent children from getting lost (e.g., "write 4 in the blue box (...), now write 7 in the orange box, ..."). Six single-digit (4, 7, 1, 6, 5, and 9) and 4 two-digit numbers (10, 11, 13, and 16) were orally presented. Maximum score is 10 and the point is awarded even if the number is mirrored. Average time required to achieve the subtest was 180 s.

¹This socio-economic index was established in Belgium in 1998 to allocate resources within the framework of the positive discrimination. It is updated every 5 years and it is constructed from the variables "per capita income, educational attainment, unemployment, occupational and comfort level of housing." To each student corresponds an index defined by its area of residence. It is the smallest administrative unit for which socioeconomic data are available. The socio-economic index is then defined based on the average of the indices of its student population; it does not correspond directly to the area of implantation, or a measure of school performance. It allows one to rank schools on a scale of 1–20, from the lowest socio-economic index to the highest. The choice of variables, indices and formula has been approved by the Government of the French Community (de Villers and Desagher, 2011; Mejias and Schiltz, 2013).

- (3) *Comparing visual number symbols (number comparison)*: Children were asked to circle the largest of the two presented Arabic digits. A total of 12 pairs of numbers ranging from “2” to “420” were presented (Four pairs of one-digit numbers with a difference of 1 to 6; Six pairs of two-digit numbers with a difference of 1 to 50; Two pairs of three-digit numbers with a difference of 17 and 69; and see section “**Appendix A**” for the detailed version of the task). An example was given on the classroom’s blackboard with “1” and “2.” The maximum score that can be obtained in this task is 12. Average time required to achieve the subtest was 210 s.
- (4) *Writing number symbols resulting from the counting of visual collections (counting)*: Children were asked to count four collections of 5 to 9 elements presented in an orderly manner (e.g., bunnies presented in a line) or presented in a disorderly manner (e.g., turtles presented in a scattered manner) and write down their answers. In the latter task three different collections were presented, two comprising the same sort of animal (turtles and sharks, respectively) and one with mixed animal sorts (lions and turtles). The maximum score that can be obtained in this task is 4 and the point is awarded even if the number is mirrored. Average time required to achieve the subtest was 120 s.
- (5) *Solving basic arithmetic problems (arithmetic problem solving)*: Children were asked to resolve a maximum of simple additions presented as “houses” of 4, 5, and 6 ($N = 18$). Operators and results of the arithmetic problems ranged between (0 and 6). The maximum score that can be obtained in this task is 18 and the point is awarded even if the number is mirrored. Average time required to achieve the subtest was 6 min.

Correct answers were scored as 1, wrong answers as 0.

Classical Mathematical Tests

To assess children’s formal mathematical skills when entering primary schooling (i.e., at T1, simultaneously with the MSR test administration) and after one entire year of formal schooling (i.e., at T2, during the first month of 2nd Grade), children were given two different classical mathematical screening tests.

The arithmetic number fact test

Tempo Test Rekenen, TTR; De Vos (1992) this test consists in two lists of arithmetic number fact problems, consisting of additions and subtractions, respectively. Children have to solve as many operations as possible within 1 min per condition. There are enough operations planned so that the child does not reach the end of the test in 1 min. Correct answers were scored as 1, wrong answers as 0. A child’s total score in the *Arithmetic Number Fact Test* corresponded to the sum of the scores obtained in the respective tasks. The TTR test was administered at T1 and T2.

The kortrijk arithmetic test

Kortrijkse Rekentest-Revisie, KRT-R; Baudonck et al. (2006) this standardized test measures children’s mathematical abilities through two subscales, these subscales correspond to the mental arithmetic computation (e.g., $43 + 36 = \dots$) and the number

system knowledge (e.g., 99 comes just after ...) and are both scored on a maximum of 30 points. Correct answers were scored as 1, wrong answers as 0. There is no time limit to accomplish the test. The maximum score that can be obtained on this scale is 60. The KRT test was administered at T2 only as children need several months of formal schooling before this test can be administered.

STATISTICAL ANALYSIS

The entire sample size could not be included in the following consistency analyses due to partial loss of data describing participants’ performance in each item. Accordingly, they were based on 158 out of 346 participants. The entire sample size is included in the other analyses. Collected data is available in the **Supplementary Material**.

Reliability

Reliability was measured by assessing internal consistency for each of the five tasks (number identification, number writing, number comparison, counting, and arithmetic problem solving) through Cronbach’s alpha and corrected item-total correlations. The corrected item-total correlation is the correlation of a selected item in one dimension with the other remaining items of that dimension. The impact of items on internal consistency was assessed by using Cronbach’s alpha with one-at-a-time deletion procedure. Cronbach’s alpha is expected to exceed 0.7 (Nunnally and Bernstein, 1978). We will consider this criterion as satisfied if 95% confidence intervals touch 0.7. Should a task not withstand the criterion, it will be excluded from further analyses.

Validity

We evaluated construct, convergent and criterion validity. Construct validity refers to the degree to which a test measures what it claims to be measuring. Convergent validity is the degree to which measures of constructs that should theoretically be related, are in fact related. Criterion validity is the extent to which a test result can be used to predict the outcome of interest.

Construct Validity

We assume that the competence underlying performance on the MSR test is essentially unidimensional and can thus be summarized in one total score. We evaluated this assumption by looking at the interrelation of all psychometrically valid MSR tasks. We expected the Pearson correlation coefficients to be positive and significant. Unidimensionality was further assessed with principal component analysis (PCA). We expected that, adhering to the Kaiser (1960) criterion (keep only components with an Eigenvalue above 1), only 1 factor would be retained.

Convergent Validity

An overall performance score for the MSR test was computed as the average of the POMP (percent of maximum performance) of all psychometrically valid tasks. In order to assess convergent validity, we computed Pearson correlations of this score with the classical mathematical tests (CMAS), composed of TTR proposed at T1 and T2, and the KRT proposed at T2.

Criterion Validity

As individuals achieving below the 25th percentile in standardized mathematical tests are classically identified as *low math achievers* (Geary, 2011), we can suppose that children in this lower quartile are at risk of developing low mathematical abilities. Moreover, children achieving at the lowest 7% should be at risk of *specific learning disorder in math*.

We created one combined indicator for students' mathematical ability at T2 (combined mathematical ability score; CMAS): TTR at T2 and KRT scores were both standardized and a sum score was created of these 2. Based on this sum score and the thresholds mentioned above, students were classified as "not at risk" (performance above the 25th percentile), "low math achievers" (7th percentile – equal to or below 25th percentile), or "potential specific learning disorder in math" (equal to or below 7th percentile).

For these groups we compared mean MSR scores using a one-way ANOVA and a *post hoc* Tukey test, expecting to find that scores would tend to be gradually lower from students "not at risk" over "low math achievers" to "potential specific learning disorder in math." Additionally, using cross tables, we compared the probability of identifying students classified as "potential specific learning disorder in math" during T2 with our MSR test to the probability of identifying them with the TTR at T1, using the 25th percentile criterion.

Using multiple linear regression, we finally checked whether students' performance on the MSR test explained variance in their score on the KRT at T2 over and above that explained by their performance on the TTR at T1.

Statistical analyses were performed using RStudio version 1.0.136.

RESULTS

Item difficulty ranges, mean POMP scores per task as well as the POMP score ranges can be found in **Table 1**.

Reliability

The Cronbach's alpha coefficients for the five tasks are reported in **Table 1** and range from 0.11 to 0.95. Due to low internal consistencies of the tasks "number identification" and

"counting," these two tasks were dropped from further analyses. For the three remaining tasks, corrected item-total correlations coefficients were all $r(156) \geq 0.25$, $p < 0.01$.

Validity

Construct Validity

Pearson correlations between the three remaining tasks were all positive and highly significant (see **Figure 1**). Effect sizes ranged from medium $r(344) = 0.28$ to large $r(344) = 0.49$, with $p < 0.001$ for each correlation coefficient.

The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.62, so above the commonly recommended value of 0.6, and Bartlett's test of sphericity was significant ($\chi^2(3) = 147.34$, $p < 0.001$). Communalities were all well above 0.3. Given these indicators, PCA was deemed to be suitable. Eigenvalues of the extracted components were 1.76, 0.74, and 0.50, with the first factor explaining 59% of the total variance. As expected, only one factor is to be retained according to the Kaiser criterion.

Convergent Validity

Since PCA confirmed that one factor explains the majority of the variance in task performance for number writing, number comparison and arithmetic problem solving, an overall score in the MSR test was computed as the average of the POMP of the three tasks. The mean score of the test ($n = 346$) was $M = 0.74$, $SD = 0.26$, with a minimum equal to 0.03 and a maximum of 1. Distribution of the score is depicted in **Figure 2**.

The MSR score significantly correlated with TTR at T1 $r(344) = 0.57$, $p < 0.001$ TTR at T2 $r(344) = 0.51$, $p < 0.001$ and KRT at T2 $r(344) = 0.51$, $p < 0.001$, see **Figure 3**.

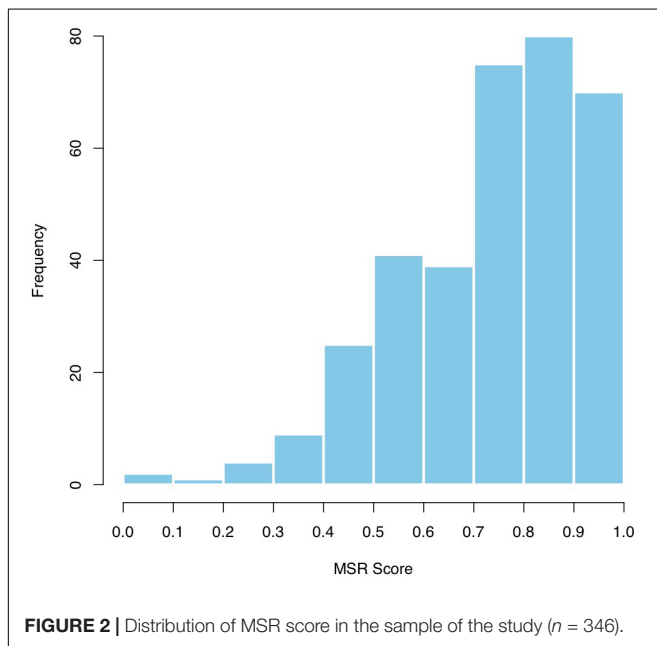
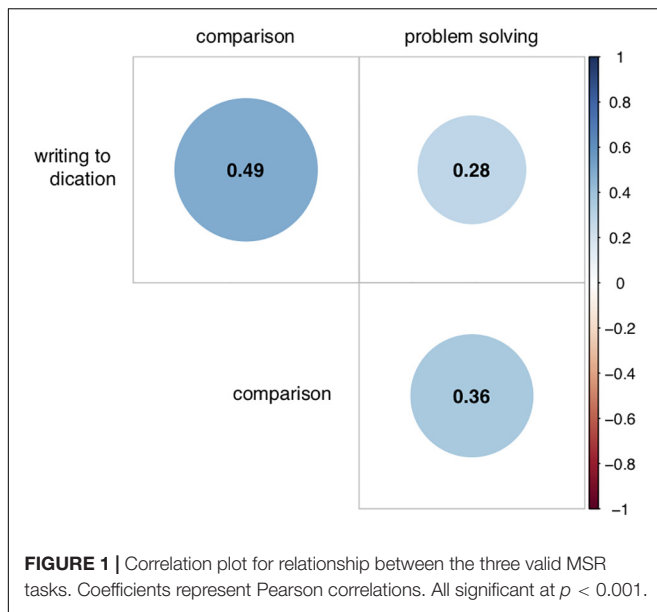
Criterion Validity

The distribution of the CMAS (centered and standardized) is presented in **Figure 4**. The Pearson correlation between CMAS and MSR is $r(344) = 0.56$, $p < 0.001$. The boxplots in **Figure 5** visualize the finding that, as expected, MSR scores tend to be lower for students classified as "potential specific learning disorder in math." For the students classified as "low math achievers," scores tend to be somewhat better, but still lower than for students that were identified as "not at risk." A one-way between subjects ANOVA confirms that mean scores of the 3 performance groups are significantly different with a large effect

TABLE 1 | Task performance.

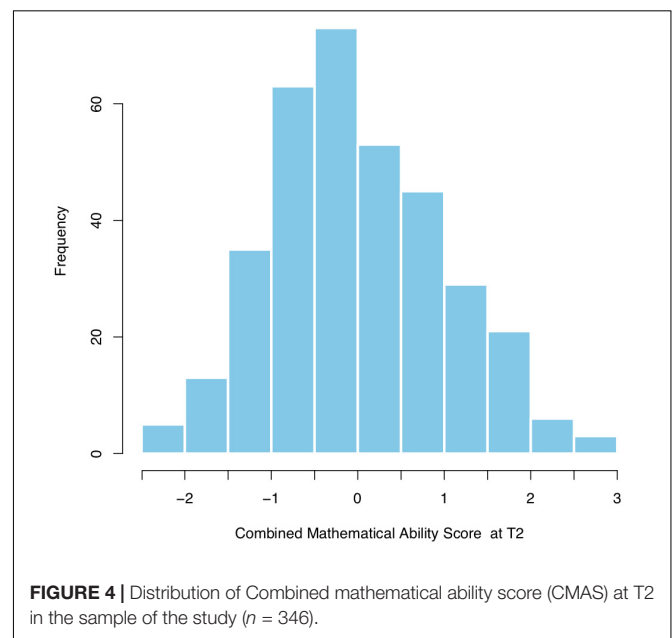
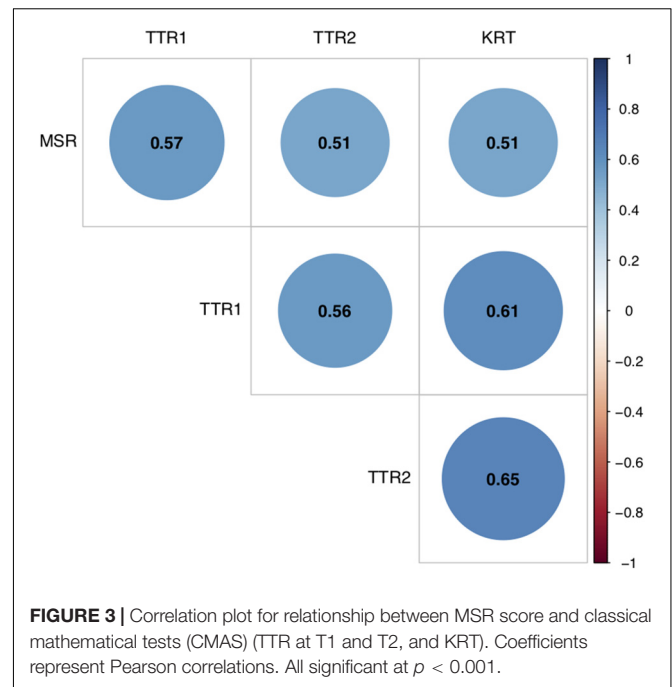
Task	N	M POMP (SD)	POMP Range		Item difficulty range	Dropped	Internal consistency	
			Theor.	Emp.			α (95% CI)	r cor.
Number identification	158	0.98 (0.06)	0–1	0.63–1	96 – 100	3, 6, 8	0.53 (0.43 – 0.64)	–0.031 – 0.59
Number writing	158	0.87 (0.13)	0–1	0.4–1	38 – 100	1, 4, 5	0.63 (0.55 – 0.75)	0.35 – 0.60
Number comparison	158	0.74 (0.19)	0–1	0.08–1	36 – 96	/	0.70 (0.64 – 0.77)	0.25 – 0.68
Counting	158	0.95 (0.12)	0–1	0.5–1	91 – 97	/	0.11 (–0.11 – 0.34)	–0.07 – 0.53
Arithmetic problem solving	158	0.66 (35)	0–1	0–1	50 – 82	/	0.95 (0.94 – 0.96)	0.56 – 0.82

POMP, Percentage of maximum performance; CI, confidence interval; Dropped, items that were dropped from internal consistency analysis due to lack of variance (all answered correctly).



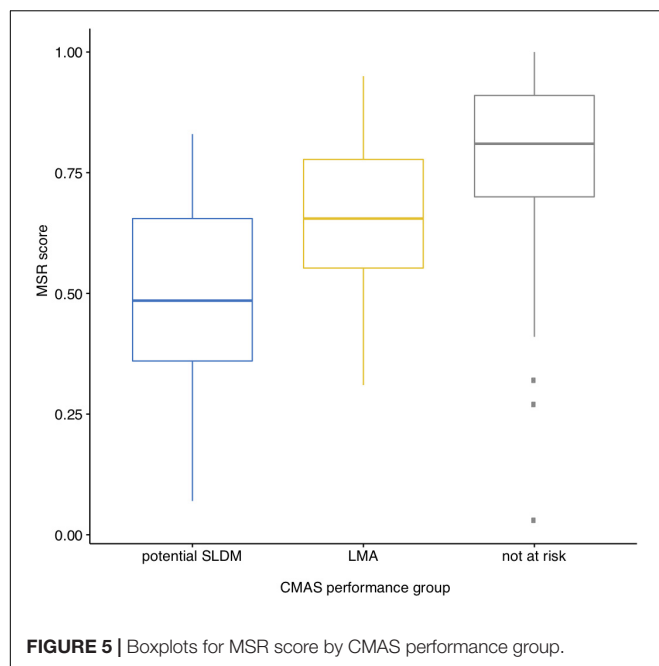
size $F(2, 343) = 50.94$, $p < 0.001$, $\eta_p^2 = 0.23$. A *post hoc* Tukey test showed that, as expected, “low math achievers” ($M = 0.65$, $SD = 0.15$) performed significantly better ($p < 0.001$) than students with a “potential learning disorder in math” ($M = 0.50$, $SD = 0.19$), but significantly worse ($p < 0.001$) than students “not at risk” ($M = 0.79$, $SD = 0.16$).

Table 2 presents the cross tabulation for performance grouping at T1 (with NRS and TTR) and T2 (CMAS). Using the MSR, 69% of students (compared to 66% with the TTR) classified as “potential specific learning disorder in math” at T2 were identified at least as “low math achievers” at T1 and 42% (compared to 31%) were already identified as “potential specific learning disorder in math.” Only 3% of students identified as



“potential specific learning disorder in math” with the MSR were later classified as “not at risk” (compared to 7% with the TTR).

Multiple regression was used in order to determine whether the MSR score would explain variance within CMAS performance over and above the variance explained by the TTR score at T1 ($CMAS \sim TTR \text{ T1} + MSR$). Results indicate that together MSR and TTR at T1 explain 41% of the variance within CMAS at T2 ($R^2 = 0.47$, $F(2, 343) = 151.5$, $p < 0.001$) with highly significant contributions from both indicators ($\beta_{TTR} = 0.47$, $p < 0.001$ and $\beta_{MSR} = 0.30$, $p < 0.001$).



Performance on the MSR test thus explains additional variance in mathematical ability on T2.

DISCUSSION

Children's academic level at school entrance, i.e., their school readiness, is very important for their future academic success and professional career (Duncan et al., 2007). Detailed knowledge about children's early abilities allows optimal adaptation of learning and instruction to their individual needs. It is therefore critical to accurately and efficiently assess school starter's abilities in the core domains of schooling, such as mathematics.

The present study aimed to design a test that allows teachers or any professional working with children (e.g., school psychologists, speech therapists, school doctors) to assess young children's MSR when they enter formal schooling in a simple, rapid and efficient manner. Such a MSR test should provide

insights into children's numerical abilities at the beginning of the 1st Grade by revealing their strengths and/or weaknesses, thereby allowing for the anticipation of their later achievements and/or problems in mathematics. The test aims to differentiate between children with distinct math ability levels, focusing in particular on the identification of children with performances in the lower range. Importantly, it is not a neuro-psychological test battery allowing full-fledged diagnosis but the test aims to inform teachers and interested professionals about children's early mathematics skills to guide their future educational set-up and/or orientation toward specific diagnosis and care measures on a solid evidence basis.

The tasks included in the test systematically related to theories of neuro-cognitive development as well as to academic competence standards, thereby ensuring that children's early mathematical abilities are measured in a cognitively accurate and valid manner. In addition, they are easy to use and can be readily interpreted by teachers. The initial test version included 5 tasks assessing children's mastery of visual number symbols: identifying visual number symbols, writing numbers to dictation, comparing visual number symbols, writing number symbols resulting from the counting of visual collections, as well as solving basic arithmetic problems. After carefully assessing the internal consistency of the different tasks, the final and validated test version retained three tasks: writing numbers to dictation, comparing visual number symbols and solving basic arithmetic problems. Internal consistency indeed indicated that the tasks consisting in identifying visual number symbols and in writing number symbols following the counting of visual collections needed to be excluded. Those two tasks lacked sensitivity and demonstrated a very low internal consistency. At the first stage of test construction, it seemed important to include the number identification task as it corresponds to a basic entry-level skill, preceding the ability to read and understand visual symbolic numbers. The same applies to the counting task, in which visual collections that had to be counted consisted of 5 to 9 elements. These type of tasks have been used in well-known diagnostic tests such as the TEDI-Math (Van Nieuwenhoven et al., 2008). The TEDI-Math is used for diagnosis of numerical learning disorders from the end of the 2nd year of kindergarten to the end of 3rd Grade of elementary school. Yet, the fact that the number

TABLE 2 | Cross table for CMAS performance groups with MSR and TTR T1 performance groups.

		MSR and TTR T1 performance groups						Row total
		Potential specific learning disorder in math		low math achievers		Not at risk		
		MSR	TTR1	MSR	TTR T1	MSR	TTR T1	
CMAS performance groups	Potential specific learning disorder in math	11 (42)	8 (31)	7 (27)	9 (35)	8 (31)	9 (35)	26
	low math achievers	9 (15)	10 (16)	19 (31)	19 (31)	34 (55)	33 (53)	62
	Not at risk	8 (3)	17 (7)	35 (14)	27 (11)	215 (83)	214 (83)	258
	Column total	28	35	61	55	257	256	346

Values represent counts with row percentages in brackets.

identification task was not a sensitive measure at the beginning of 1st Grade is not surprising considering that children in kindergarten (from 4-to-6-year-olds) revealed remarkably good knowledge of visual number symbols. They are thus able to estimate, compare, add and subtract 2-digit numbers, based on their approximate number sense (Gilmore et al., 2007; Mejias and Schiltz, 2013). Concerning the present counting task, it appeared to be much simpler than the task proposed in the TEDI-Math, in which the child must provide an answer based on a detailed understanding of elaborate language instructions.

Considering the final version of the MSR task comprising the three tasks “number writing,” “number comparison,” and “arithmetic problem solving” a PCA indicated that the final test can be characterized by a single dimension involving basic number skills. The three internally consistent tasks of the MSR test (as all corrected item-total correlations were greater than 0.25) were not redundant, indicating that all subtests contribute to the measure of early mathematical abilities. Performance on all three tasks thus contributed relevant information to explaining individual differences in early mathematical abilities, which are considered to be essential scaffolds for later formal arithmetical abilities. It was indeed proposed that mathematical abilities develop quasi-hierarchically, with more mature and complex mathematical knowledge building up on more basic skills (Claessens et al., 2009; Jordan et al., 2009; Claessens and Engel, 2013; Watts et al., 2014; Aunio and Räsänen, 2016). The test thus notifies about the mastery of visual symbols, by providing information about children’s abilities to write numbers to dictation (i.e., referring to transcoding abilities; e.g., Molfese et al., 2006), to compare Arabic digits (i.e., referring to number magnitude representations and place-value understanding; e.g., Nosworthy et al., 2013; Brankaer et al., 2017) and to solve basic arithmetic problems (i.e., referring to basic computational skills; e.g., Jordan et al., 2006).

The scores in the MSR test were distributed over the entire performance range, going from very low (0.03) to perfect (1.0). This indicates that the difficulty level of the test is well adapted to capture the performance of all children attending 1st Grade. Critically, children’s performances on the MSR test at school entrance predicted their mathematical performances 1 year later, yielding a correlation of 0.56 with a combined measure of two CMAS. Moreover, the 42% of children, who were identified as “potential specific learning disorder in math” since they scored below 7% in the MSR test (i.e., 3% of the total group), also achieved below 7% 1 year later in 2nd Grade. In comparison, only 31% of the children were accurately classified based on the TTR1. The present test therefore allows to anticipate later mathematics achievements, which in turn facilitates early actions specifically adapted to a child’s profile. Especially, those children scoring below 7% run the risk of significantly falling behind if no specifically dedicated measures are taken. They should therefore be oriented toward further psychological support and special needs education, if the specific learning disorder in math is confirmed by classical neuro-psychological test procedures. Nevertheless, a certain number of classification errors can arise with the novel MSR test, as with the more established tests TTR and KRT. These might reflect for instance math problems

arising after the first assessment point T1, or measurement noise occurring at T1 or T2 and which leads to performances that do not truly reflect children mathematical abilities (i.e., tiredness, lack of concentration or lack of motivation during test taking).

In line with previous studies in preschoolers (Duncan et al., 2007; Aunio and Niemivirta, 2010; Pagani et al., 2010; Romano et al., 2010) the present results confirm that early numeracy performances are a good predictor of later more elaborate math performance. Mastery of the Arabic number system is a major challenge in math skill acquisition, as it emerges from the progressive association of numerical meaning to visual symbols, which takes place over a 2–3 year period from age 3 onward (see Thevenot and Fayol, 2018, for a review). This corresponds to the 3rd and 4th stage of the von Aster and Shalev (2007; see also Kaufmann et al., 2014), referring to the mastery of Arabic number representation and their ordering on a mental number line, respectively. From a formal point of view, these acquisitions classically emerge through explicit academic learning during kindergarten and are subsequently reinforced in primary school. Since approximately 30 years (with the widespread use of digital displays), implicit learning of visual number symbols also often occurs in children’s home environment (Thevenot and Fayol, 2018). The more children are exposed to informal learning opportunities at home, the better they perform on basic number skill tasks (Melhuish et al., 2008; Benavides-Varela et al., 2016). Accordingly, some of the children scoring below 25% may be those lacking informal number activities, therefore lagging behind peers, who experience more informal numerical activities in their early home environment (see Ramani and Siegler, 2014, for a review). Activities that include Arabic numbers have been shown to help these children overcome their gap (e.g., Ramani and Siegler, 2008; Siegler and Ramani, 2008). Apart from lacking numerical stimulation, some children scoring below 7% may additionally suffer from specific learning disorder in math, therefore requiring even more targeted follow-ups. The MSR does not allow disentangling these two problem sources. Yet in either case, it is important to be able to efficiently and reliably identify children as soon as possible.

As opposed to existing tools (i.e., exhaustive test batteries, minimal screeners), our MSR test aims to assess children’s early mathematical abilities, while being easy to administer as well as readily interpretable by any early childhood professional such as teachers, school psychologists, speech therapists, and school doctors. The present test version is, however, limited to (pre)-school curricula including explicit instruction of number symbols up to 10 and using French as instruction language. In future studies norms should be collected in French-speaking countries with similar curricula. Furthermore, it will be important to address potential performance influences due to children’s socio-economic and (multi-)lingual environment (e.g., Mejias and Schiltz, 2013; Van Rinsveld et al., 2015).

In sum, the MSR test offers a tool that is short (approximately 15 min), can be administered individually or collectively in the classroom setting and allows to reliably evaluate early mathematic abilities, encompassing writing numbers to dictation, comparing visual number symbols, and solving basic arithmetic problems. It complements the existing studies based on math school

readiness of preschool children (Blair and Razza, 2007; Duncan et al., 2007; Pagani et al., 2010), by providing a test that can be administered by those teachers and/or health professionals that are accompanying the children throughout the two first years of elementary school. Since the MSR test has proven to be an efficient predictor of children's proficiency in classical math tests administered 1 year later, it can be used to detect children who are at risk of performing low in mathematics. Empirically validated curricula and specialized neuro-psychological diagnostics and interventions can then be applied depending on the child's ability level (Wilson et al., 2006; Butterworth and Laurillard, 2010; Clements and Sarama, 2011; Kucian et al., 2011; Skwarchuk et al., 2014; Iuculano et al., 2015).

ETHICS STATEMENT

The present research has obtained the full consent of the Institutional Ethics Committee of the Université Catholique de Louvain (Belgium).

AUTHOR CONTRIBUTIONS

SM and CS conceived the test and wrote the manuscript. SM carried out the experiments. CM performed the analysis and

designed the figures. All authors discussed the results and contributed to the final manuscript.

FUNDING

This study received the support of the Graduate school of teaching and Education Lille Nord de France (ESPE LNF).

ACKNOWLEDGMENTS

We thank Nathanael Larigaldie (Durham University), Adrien Lecoivre (CHU de Lille), and Genin Michaël (Université de Lille) for their assistance in data analyses.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01173/full#supplementary-material>

DATA SHEET S1 | All novel test material used and data collected in the present study in order to build the "mathematical school readiness test".

APPENDIX A | Initial version of the "mathematical school readiness test"; only 3 items are included in the final version of the test.

REFERENCES

- American Psychiatric Association [APA] (2013). *Specific Learning Disorder. Diagnostic and Statistical Manual of mental disorders (DSM-5®)*. Washington, DC: American Psychiatric Association.
- Aunio, P., and Niemivirta, M. (2010). Predicting children's mathematical performance in grade one by early numeracy. *Learn. Individ. Differ.* 20, 427–435. doi: 10.1016/j.lindif.2010.06.003
- Aunio, P., and Räsänen, P. (2016). Core numerical skills for learning mathematics in children aged five to eight years—a working model for educators. *Eur. Early Child. Educ. Res. J.* 24, 684–704. doi: 10.1080/1350293x.2014.996424
- Baudonck, M., Debusschere, A., Dewulf, B., Samyn, F., Vercaemst, V., and Desoete, A. (2006). *Kortrijkse Rekestest Revisie (KRT-R)*. Kortrijk: Revalidatiecentrum Overleie.
- Benavides-Varela, S., Butterworth, B., Burgio, F., Arcara, G., Lucangeli, D., and Semenza, C. (2016). Numerical activities and information learned at home link to the exact numeracy skills in 5-6 years-old children. *Front. Psychol.* 7:94. doi: 10.3389/fpsyg.2016.00094
- Blair, C., and Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Dev.* 78, 647–663. doi: 10.1111/j.1467-8624.2007.01019.x
- Brankaer, C., Ghesquière, P., and De Smedt, B. (2017). Symbolic magnitude processing in elementary school children: a group administered paper-and-pencil measure (SYMP Test). *Behav. Res. Methods* 49, 1361–1373. doi: 10.3758/s13428-016-0792-3
- Butterworth, B., and Laurillard, D. (2010). Low numeracy and dyscalculia: identification and intervention. *ZDM* 42, 527–539. doi: 10.1007/s11858-010-0267-4
- Claessens, A., and Engel, M. (2013). How important is where you start? Early mathematics knowledge and later school success. *Teachers Coll. Rec.* 115:n6.
- Claessens, A., Duncan, G., and Engel, M. (2009). Kindergarten skills and fifth-grade achievement: evidence from the ECLS-K. *Econ. Educ. Rev.* 28, 415–427. doi: 10.1016/j.econedurev.2008.09.003
- Clements, D. H., and Sarama, J. (2011). Early childhood mathematics intervention. *Science* 333, 968–970. doi: 10.1126/science.1204537
- Currie, J., and Thomas, D. (1999). *Early Test Scores, Socioeconomic Status and Future Outcomes*. Cambridge, MA: National bureau of economic research, w6943.
- de Villers, J., and Desagher, C. (2011). *L'indice Socio-Economique Des Écoles. COMMENT ça Marche? A Quoi ça Sert?*. Tallahassee, FL: Les analyses de la FAPEO.
- De Vos, T. (1992). *Tempo Test Rekenen (TTR)*. Nijmegen: Berkhout.
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-n
- Do, C.-L. (2007). "Les représentations de la grande difficulté scolaire par les enseignants," in (*d. l. E. s. e. d. l. R. Ministère de l'Éducation Nationale, Direction de l'évaluation, de la Prospective et de la Performance*, Vol. 182, ed. Centre de documentation de la DEPP (Paris: Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche).
- Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Dev. Sci.* 11, 650–654. doi: 10.1111/j.1467-7687.2008.00713.x
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Dev. Psychol.* 43, 1428–1446. doi: 10.1037/0012-1649.43.6.1428
- Feigenson, L., Dehaene, S., and Spelke, E. S. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307–314. doi: 10.1016/j.tics.2004.05.002
- Fischbach, A., Budson, T. G., Preckel, F., Martin, R., and Brunner, M. (2013). Do teacher judgments of student intelligence predict life outcomes? *Learn. Individ. Differ.* 27, 109–119. doi: 10.1016/j.lindif.2013.07.004
- Geary, D. C. (2011). Consequences, characteristics, and causes of mathematical learning disabilities and persistent low achievement in mathematics. *J. Dev. Behav. Pediatr.* 32:250. doi: 10.1097/DBP.0b013e318209edef

- Geary, D. C., Hoard, M. K., and Hamson, C. O. (1999). Numerical and arithmetical cognition: patterns of functions and deficits in children at risk for a mathematical disability. *J. Exp. Child Psychol.* 74, 213–239. doi: 10.1006/jecp.1999.2515
- Gilmore, C. K., McCarthy, S. E., and Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature* 447, 589–591. doi: 10.1038/nature05850
- Gross-Tsur, V., Manor, O., and Shalev, R. S. (1996). Developmental dyscalculia: prevalence and demographic features. *Dev. Med. Child Neurol.* 38, 25–33. doi: 10.1111/j.1469-8749.1996.tb15029.x
- Halberda, J., Mazzocco, M. M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* 455, 665–668. doi: 10.1038/nature07246
- Iuculano, T., Rosenberg-Lee, M., Richardson, J., Tenison, C., Fuchs, L., Supekar, K., et al. (2015). Cognitive tutoring induces widespread neuroplasticity and remediates brain function in children with mathematical learning disabilities. *Nat. Commun.* 6:8453.
- Iuculano, T., Tang, J., Hall, C. W. B., and Butterworth, B. (2008). Core information processing deficits in developmental dyscalculia and low numeracy. *Dev. Sci.* 11, 669–680. doi: 10.1111/j.1467-7687.2008.00716.x
- Jordan, N. C., Kaplan, D., Locuniak, M. N., and Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learn. Disabil. Res. Pract.* 22, 36–46. doi: 10.1111/j.1540-5826.2007.00229.x
- Jordan, N. C., Kaplan, D., Oláh, L. N., and Locuniak, M. N. (2006). Number sense growth in kindergarten: a longitudinal investigation of children at risk for mathematics difficulties. *Child Dev.* 77, 153–175. doi: 10.1111/j.1467-8624.2006.00862.x
- Jordan, N. C., Kaplan, D., Ramineni, C., and Locuniak, M. N. (2009). Early math matters: kindergarten number competence and later mathematics outcomes. *Dev. Psychol.* 45, 850–867. doi: 10.1037/a0014939
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* 20, 141–151. doi: 10.1177/001316446002000116
- Kaufmann, L., Kucian, K., and von Aster, M. (2014). “Development of the numerical brain,” in *The Oxford Handbook of Numerical Cognition*, eds R. C. Kadosh, and A. Dowker (Oxford: Oxford University Press). doi: 10.1093/oxfordhb/9780199642342.013.008
- Koontz, K. L., and Berch, D. B. (1996). Identifying simple numerical stimuli: processing inefficiencies exhibited by arithmetic learning disabled children. *Math. Cogn.* 2, 1–23.
- Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., et al. (2011). Mental number line training in children with developmental dyscalculia. *NeuroImage* 57, 782–795. doi: 10.1016/j.neuroimage.2011.01.070
- Lafay, A., and Helloin, M. C. (2016). *Examath 8-15, Batterie Informatisée D'examen Des Habiletés Mathématiques*. Grenade: HappyNeuron.
- Landerl, K., Bevan, A., and Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8-9-year-old students. *Cognition* 93, 99–125. doi: 10.1016/s0010-0277(04)00014-9
- Landerl, K., Fussenegger, B., Moll, K., and Willburger, E. (2009). Dyslexia and dyscalculia: two learning disorders with different cognitive profiles. *J. Exp. Child Psychol.* 103, 309–324. doi: 10.1016/j.jecp.2009.03.006
- Landerl, K., and Kölle, C. (2009). Typical and atypical development of basic numerical skills in elementary school. *J. Exp. Child Psychol.* 103, 546–565. doi: 10.1016/j.jecp.2008.12.006
- Mejias, S., and Schiltz, C. (2013). Estimation abilities of large numerosities in kindergartners. *Front. Psychol.* 4:518. doi: 10.3389/fpsyg.2013.00518
- Melhuish, E. C., Sylva, K., Sammons, P., Siraj-Blatchford, I., Taggart, B., Phan, M., et al. (2008). Preschool influences on mathematics achievement. *Science* 321, 1161–1162. doi: 10.1126/science.1158808
- Mix, K. S., Prather, R. W., Smith, L. B., and Stockton, J. D. (2014). Young children's interpretation of multidigit number names: from emerging competence to mastery. *Child Dev.* 85, 1306–1319. doi: 10.1111/cdev.12197
- Molfese, V. J., Beswick, J., Molnar, A., and Jacobi-Vessels, J. (2006). Alphabetic skills in preschool: a preliminary study of letter naming and letter writing. *Dev. Neuropsychol.* 29, 5–19. doi: 10.1207/s15326942dn2901_2
- Mussolin, C., Mejias, S., and Noël, M.-P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition* 115, 10–25. doi: 10.1016/j.cognition.2009.10.006
- National Research Council (2001). *Adding it up: Helping Children Learn Mathematics*. Washington, DC: National Academies Press.
- Nosworthy, N., Bugden, S., Archibald, L., Evans, B., and Ansari, D. (2013). A two-minute paper-and-pencil test of symbolic and nonsymbolic numerical magnitude processing explains variability in primary school children's arithmetic competence. *PLoS One* 8:e67918. doi: 10.1371/journal.pone.0067918
- Nunnally, J. C., and Bernstein, I. H. (1978). *Psychometric Theory*. New York, NY: McGraw-Hill.
- Pagani, L. S., Fitzpatrick, C., Archambault, I., and Janosz, M. (2010). School readiness and later achievement: a french canadian replication and extension. *Dev. Psychol.* 46, 984–994. doi: 10.1037/a0018881
- Ramani, G. B., and Siegler, R. S. (2008). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Dev.* 79, 375–394. doi: 10.1111/j.1467-8624.2007.01131.x
- Ramani, G. B., and Siegler, R. S. (2014). “How informal learning activities can promote children's numerical knowledge,” in *The Oxford Handbook of Numerical Cognition*, eds R. C. Kadosh and A. Dowker (Oxford: Oxford University Press).
- Reigosa-Crespo, V., Valdés-Sosa, M., Butterworth, B., Estévez, N., Rodríguez, M., Santos, E., et al. (2012). Basic numerical capacities and prevalence of developmental dyscalculia: the havana survey. *Dev. Psychol.* 48, 123–135. doi: 10.1037/a0025356
- Rivera-Batiz, F. L. (1992). Quantitative literacy and the likelihood of employment among young adults in the United States. *J. Hum. Res.* 27, 313–328.
- Romano, E., Babchishin, L., Pagani, L. S., and Kohen, D. (2010). School readiness and later achievement: replication and extension using a nationwide canadian survey. *Dev. Psychol.* 46, 995–1007. doi: 10.1037/a0018880
- Rousselle, L., and Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: a comparison of symbolic vs non-symbolic number magnitude processing. *Cognition* 102, 361–395. doi: 10.1016/j.cognition.2006.01.005
- Rubinsten, O., and Henik, A. (2005). Automatic activation of internal magnitudes: a study of developmental dyscalculia. *Neuropsychology* 19, 641–648. doi: 10.1037/0894-4105.19.5.641
- Shalev, R. S., Manor, O., and Gross-Tsur, V. (2005). Developmental dyscalculia: a prospective six-year follow-up. *Dev. Med. Child Neurol.* 47, 121–125. doi: 10.1111/j.1469-8749.2005.tb01100.x
- Siegler, R. S., and Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Dev. Sci.* 11, 655–661. doi: 10.1111/j.1467-7687.2008.00714.x
- Skwarchuk, S. L., Sowinski, C., and LeFevre, J. A. (2014). Formal and informal home learning activities in relation to children's early numeracy and literacy skills: the development of a home numeracy model. *J. Exp. Child Psychol.* 121, 63–84. doi: 10.1016/j.jecp.2013.11.006
- Thevenot, C., and Fayol, M. (2018). *Les Activités numériques : Des Premières Intuitions à la Résolution de problèmes Dessiner, Lire, Écrire, Calculer: un Regard Neuf*. Paris: Editions in press, 245–315.
- Van Lint, S. (2010). *L'entrée Dans les Mathématiques à L'école Maternelle*. Bruxelles: Ministère de la Communauté Française.
- Van Nieuwenhoven, C., Grégoire, J., and Noël, M.-P. (2008). *TEDI-Math: Test Diagnostique des Compétences de Base en Mathématiques: Manuel*. Paris: du Centre de psychologie appliquée.
- Van Rinsveld, A., Brunner, M., Landerl, K., Schiltz, C., and Ugen, S. (2015). The relation between language and arithmetic in bilinguals: insights from different stages of language acquisition. *Front. Psychol.* 6:265. doi: 10.3389/fpsyg.2015.00265
- von Aster, M. G. (2000). Developmental cognitive neuropsychology of number processing and calculation: varieties of developmental dyscalculia. *Eur. Child Adolesc. Psychiatry* 9, 41–57.
- von Aster, M. G., and Shalev, R. S. (2007). Number development and developmental dyscalculia. *Dev. Med. Child Neurol.* 49, 868–873. doi: 10.1111/j.1469-8749.2007.00868.x

- Watts, T. W., Duncan, G. J., Siegler, R. S., and Davis-Kean, P. E. (2014). What's past is prologue: relations between early mathematics knowledge and high school achievement. *Educ. Res.* 43, 352–360. doi: 10.3102/0013189X14553660
- Willburger, E., Fussenegger, B., Moll, K., Wood, G., and Landerl, K. (2008). Naming speed in dyslexia and dyscalculia. *Learn. Individ. Differ.* 18, 224–236. doi: 10.1016/j.lindif.2008.01.003
- Wilson, A. J., Revkin, S. K., Cohen, L., Cohen, D., and Dehaene, S. (2006). An open trial assessment of “The number race”, an adaptive computer game for remediation of dyscalculia. *Behav. Brain Funct.* 2:20.

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

Copyright © 2019 Mejias, Muller and Schiltz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership